SEEDS Standards and Interface Processes Draft Survey Report

Draft Version 1.10

September 18, 2002

EXECUTIVE SUMMARY

This report makes the basic assumption that use of well-defined interfaces and standards is desirable in a distributed heterogeneous setting as described in the various ESE SEEDS vision documents. The validity of that assumption will be the subject of other work and reports. This report provides some answers to the following two questions:

1. How were standards adopted within ESE in the past and how successful was the adoption in terms of actual implementation experience?
2. What are some of the formal standards bodies that produce standards relevant to ESE, what are their internal processes, and how can ESE benefit from participating in these standards bodies, particularly in light of the experiences captured in the first part?

Both questions are addressed based on a sampling of experiences, both from the contributors own experiences as well as from interviews conducted with ESE staff and affiliated members. A set of lessons learned has been distilled primarily from the answers to the first question and is presented here. The results are not surprising. The basic principles behind successful selection, adoption, and implementation of interfaces and standards within ESE are the suitability of a particular interface to the problem at hand, its simplicity, the availability of tools to help projects employ it, and some education and evangelization about the benefits of standards to the organization.

CONTRIBUTORS

Study team:

Kenneth R. McDonald, NASA Goddard Space Flight Center *(Study Team Lead)*
Jean-Jacques Bedet, Science Systems & Applications, Inc. (SSAI)
Helen Conover, University of Alabama @ Huntsville
Allan Doyle, International Interfaces, Inc.
Yonsook Enloe, Stinger Ghaffarian Technologies (SGT)
Dr. John D. Evans, Global Science & Technology, Inc. (GST)
Dr. Ramachandran Suresh, Mayur Technologies

Consultants:

Prof. Liping Di, George Mason Uuniversity
Prof. Jim Frew, UCSB
Douglas Nebert, FGDC
Prof. Silvia Nittel, University of Maine
George Percivall, Global Science & Technology, Inc. (GST)
Lola Olsen, NASA Goddard Space Flight Center
    Wrote portions of Section 2.2.1
Dr. Don Sawyer, NASA Goddard Space Flight Center

1 INTRODUCTION

1.1 Earth Science Enterprise Background

The NASA Earth Science Enterprise (ESE) has invested in a series of Earth science information systems in the last 20 years. In the 1980s, it funded the development of a series of independent, project oriented data systems. Each project or mission managed the development, maintenance, and operations of a data system for its particular project or mission data. Although many of the projects had similar requirements for data system capabilities and similar data, data system development efforts were typically carried out in isolation and were not required to leverage or reuse across projects, nor were they required to be interoperable[1] across projects or with outside entities. As a result, reuse or leveraging of existing systems or subsystems was minimal. Also, the independent and heterogeneous systems and standards proved to be an obstacle to interdisciplinary studies and Earth system science where investigators wished to access and use data products from more than one source.

In the early 1990s, the Version 0 (V0) system was developed as an early prototype of EOSDIS. The EOSDIS V0 Information Management Systems (IMS) provided an interoperable inventory layer over existing, independent data systems to support the search and order access to heritage data collections held by NASA's Distributed Active Archive Centers (DAACs). The full EOSDIS Core System (ECS), developed during much of the 1990s, provided full system capability to ingest, process, archive, catalog, and distribute huge amounts of data from multiple new EOS missions scheduled for launch in the late 1990s. The EOSDIS Core System, developed by a single development team, provided a single system for all the EOS data centers. In part due to concerns that ECS was monolithic and difficult to interface to from outside of the system, in 1996 the National Research Council organized a review of the NASA EOSDIS project. The review report called for NASA to experiment with a PI-controlled, federation of diverse data providers. In 1997, NASA funded a Federation of cooperating, but autonomous Earth Science Information Partners (ESIPs –see http://www.esipfed.org), charged to work together as an experiment in self-governance and interoperability. Many of these ESIPs were themselves collaborations of Earth Science and Information Technology innovations. Nine Distributed Active Archive Centers (DAACs) joined the original 24 ESIPs to form an initial working Federation of 33 data providers. This group has continued to grow, and now includes 41 partners plus NASA.

[1] ISO TC/204 states that "Interoperability is the ability of systems to provide services to and accept services from other systems and to use the services so exchanged to enable them to operate effectively together."

A Program Formulation team entitled Strategies for Evolution of ESE Data Systems (SEEDS) is studying different approaches towards management of the evolution of Earth science data systems in the next two to ten years. Distributed development by multiple partners of a set of heterogeneous open systems is expected in the SEEDS era. A key question is “How can this kind of development be facilitated and managed, and how can interoperability among these heterogeneous systems be guaranteed?” One element of managing the development of Earth science data systems is the management of standards and interfaces used within the data systems. This includes the processes of standards development, standards selection, and standards implementation.

1.2 Standards Overview

Standards are essential for achieving interoperability and efficiency in all aspects of satellite mission and data system development. The successful adoption and use of standards for a particular project could potentially reduce the cost and resources required for system engineering and development.

There are at least three fundamental aspects to the use of standards within an enterprise:

1. A selection process: how specific standards are identified as having value to a project if they are employed in part or all of the enterprise.
2. An implementation process: how selected standards are actually employed within the enterprise.
3. A development process: how new standards are developed if no suitable standards exist for use within the enterprise.

These processes affect each other and tradeoffs exist when decisions about any given process are made. For instance, if a selection process is identifying standards, there is a tradeoff between how well the standard meets a need and how hard it would be to develop a standard that might better meet that need. The selection process should also take into account how easy or hard it can be to implement use of that standard within the enterprise. Furthermore since a given enterprise itself consists of a heterogeneous collection of technical, organizational, and cultural aspects, these processes have to take those aspects into account. A key principle of SEEDS seems to be its distributed and cooperative nature. It is not anticipated that SEEDS is an environment where “one size fits all” approaches will be successful (or even tolerated by its constituents).

There are several ways for an organization to select and implement standards. The common methods include

- **develop** its own standards;
- **adopt** standards developed by others;
- **extend** standards developed by others; and
- **profile** standards developed by others.

The processes associated with the different methods of setting standards are also different. Some organizations may only use one of the standard setting methods while others may use all of them. It is likely that SEEDS will use all of the above-mentioned methods.

The selection of standards could depend on several factors:

- Previous successful use of standards by the community
- Number of projects/people using the standard
- Open and transparent standards development process
- Timeliness of the standards development process
- Software tools (open source and/or COTS) support for the standard
- Evolvability of the standard
- Suitability in terms of performance (speed, accuracy, etc.)
- Cost of access to the standard itself
- Availability of the standard to all members of the community
- Quality of the standards document (readability, consistency, etc.)

The number and type of standards required to support a mission and achieve data interoperability between the mission and other missions, and/or user communities could be large. It is difficult to forecast what types of standards are suitable for the long term, particularly beyond the next five years. Developments in new technologies could potentially affect evolution and development of standards, and therefore it is important to identify standards development, selection and implementation processes that can accommodate changes in technologies and evolve.

One may categorize standards in several ways. Along one axis, there are the kinds of things that are being standardized:

- *Content* standards describe what kinds of information must be present, but don't necessarily dictate what form that information must take
- *Encoding* standards dictate specifically what form information must take, but they don't generally dictate what particular content is being encoded.
- *Process* standards dictate how to perform a particular activity.
- *Interface* standards describe the interaction between two components.
- *Transfer* standards dictate how to move information in bulk from one location to another while preserving the semantics of the information.

One may also distinguish standards by their degree of formality:

- *De facto* standards have been adopted for use but that have not been ratified or published by a formal standards body.
- *De jure* standards have been ratified and published by a formal standards body.

One may also divide standards into their spheres of influence. These can be either adopted standards or developed standards:

- *Project* standards are used within a project

- *Enterprise* standards are used within an enterprise
- *National* standards are used within a country
- *International* standards are used internationally
- Domain standards are used within a discipline or domain

1.3 Purpose and Methodology

The purpose of this document is to study different approaches to developing interfaces and standards. We analyze how ESE and similar projects in other agencies have developed and deployed standards and interfaces; and how international standards organizations develop them today. What makes an interface standard successful? What lessons have we learned from past development efforts in ESE and elsewhere? What lessons can we learn from notable standards organizations today? What kinds of interface standards development processes facilitate distributed development, open systems, and community agreement on standards? Can this open, distributed development generate standards and interfaces to support strict mission schedules? Who are the stakeholders for the various interfaces? How to facilitate agreement and adoption of standards? How to manage evolution of the interface standards?

In order to achieve these objectives, SEEDS convened a study team of experts from different organizations working actively in standards activities. Study team members interact regularly via email and through teleconferencing. To augment participation from other organizations and to get representation from large community, the study included a team of consultants.

In this report, we surveyed several standards organizations and projects that develop standards relevant to ESE. This report does not intend to be an exhaustive list of all standards relevant to the ESE. The focus is on the standard processes rather than the standards themselves. A special emphasis of the survey was the process used for the development of standards by various organizations and projects. From this study, we identify best practices and derive some preliminary lessons learned on the standard processes.

In studying how standards and interfaces were developed or deployed by major projects by ESE and others in the past, we identify which of these standards and interfaces were the most successful and why. Similarly, we identify which international standards organizations are the most successful and why, then make some recommendations on what kind of approaches are needed for the successful adoption and deployment of standards and interfaces in SEEDS. We also examine how ESE can foster activity in those areas where standards are needed but are not yet under development.

2 NASA-ESE EXPERIENCES WITH STANDARDS AND INTERFACES

2.1 Overview

The overall goal of the Version 0 EOSDIS system was to provide a cross-data-center inventory search and order capability, layered over existing and evolving data center systems. The EOSDIS Core System (ECS), on the other hand, was intended as a comprehensive, end-to-end data system to handle ingestion, processing, cataloging, archival, and distribution of all NASA Earth Science data products. This single system must handle huge data volumes, a variety of data types, and must meet the needs of the science community (both data producers and data consumers). In order to meet the ambitious data system requirements of the EOS missions, the ECS design converged on a system of tightly coupled and rigidly controlled components. The accompanying loss in flexibility was a frequent complaint from many members of the EOS science community. It drew even stronger criticism when some of the original, driving requirements, such as cross-DAAC processing, were dropped for cost, schedule or complexity issues.

As a result, the Federation of Earth Science Information Partners (ESIPs) was conceived to give data system control back to the science community, in collaboration with information technology specialists, and to enable distributed, heterogeneous development. One of the original requirements levied on all ESIPs was to provide for interoperability within the Federation in a to-be-defined System-Wide Interface Layer (SWIL). As the ESIPs developed the Federation governance structure, a Standing Committee on Interoperability was formed to define and develop the SWIL. Initially, the committee focus was on providing for *catalog interoperability*, which would allow data from all ESIPs to be discovered through a single user interface. At the same time “clusters” of ESIPs were formed to explore various common interests, including interoperable data access technologies. These technologies are to allow software applications to access data at multiple Federation sites using a common interface. Current efforts are concentrating bringing *data access interoperability* into the SWIL.

The Federation’s challenges, especially in coordinating data products and services, are similar to those expected in the SEEDS era.

2.2 Metadata Standards

2.2.1 Directory Interchange Format (DIF)

2.2.1.1 Overview

The Directory Interchange Format (DIF) is a *de facto* metadata exchange standard for the discovery of Earth science data through the CEOS International Directory Network (IDN), of which the NASA Global Change Master Directory (GCMD) is a principal member. (cf. http://gcmd.nasa.gov/)

The Services Entry Resource File (SERF) is another metadata exchange standard, similar to the DIF, recently created for the discovery of Earth science services. A hierarchical keyword structure helps to classify data services and serves an important role in the user interface.

2.2.1.2 Process and Evolution

The directory evolved from NASA's original Master Directory, which once served both the space and Earth science communities. When data management responsibilities were divided, the GCMD's Earth scientists revamped the keyword hierarchy, modified and added fields to serve the Earth science community more directly, and added new fields that are now pertinent because of related technology upgrades. These new fields are also important to the users. The set of fields includes all those that are needed to determine if any particular data set will be useful to the searcher.

All modifications and upgrades for the system result from usage metrics; relevant advances in technology; NASA requirements; and/or evaluations and recommendations by the GCMD's Science User Working Group. Additions and modifications require a consensus of agencies participating in the IDN's Interoperability Forum, which is led and moderated by NASA's Global Change Master Directory personnel.

At one time, four different user interfaces were available to users. Based on user metrics, two of these are no longer offered. The user may choose either the keyword interface or the free-text interface. The free-text interface offers access through the Z39.50 protocol.

2.2.1.3 Tools and Documentation

Both tools and documentation are available at the GCMD web site: [http://gcmd-.gsfc.nasa.gov](http://gcmd.gsfc.nasa.gov). Tools include interactive web-based interfaces for adding or updating directory entries, or exporting metadata in a variety of formats (DIF, FGDC, XML, …)

The Global Change Master Directory offers data set and data service authoring tools to assist in writing quality descriptions that the database software will validate and accept. Current authoring tools (written in Perl) are available as DIFbuilder, modDIFbuilder, SERFbuilder, modSERFbuilder, etc. In the near future, the authoring tools will be object-oriented and written in Jython to reduce the maintenance for the suite of tools. Operations client software is also available to assist in loading the database.

Complete documentation for the XML-based, object-oriented software system is available for the MD8 server, client, and database. All ISO 9001 documentation is in place, as well as plans to assure compliance with the upcoming ISO 19115 standard for geospatial metadata.

2.2.1.4 Current Status

The GCMD contains descriptions of over 10,000 datasets from data providers within ESE and from data providers around the world. The number of users for the system is increasing yearly. The quality control that the GCMD staff exercise over the dataset descriptions make this a unique directory, unmatched in content and quality by any other existing directory.

The GCMD enforces metadata “standards” in two key areas: the DIF defines metadata attributes needed for a complete directory level description of a data set, while the controlled keywords enforced for selected metadata attributes assure consistent search results for data sets cataloged within the system.

The GCMD validates compliance by checking that (1) required fields appear, (2) their syntax is correct, and (3) controlled keywords are valid. The science keywords are enforced to assure consistent search results for data sets described within the system. There are also controlled lists of keywords (valids) available to properly describe instruments, platforms, data centers, location, project / campaign, chronostratigraphic units, and Related_URL content type. In addition, through the operations client, the personnel database for the directory is carefully maintained and updated.

The GCMD staff supports the evolution of keywords - in response to community needs. New keywords may be added to the controlled lists at the science community's request within a short time, provided the keyword rules are followed. The DIF structure has evolved with care since its initial definition, and has remained backward compatible since MD4.

2.2.2 V0 Metadata Model

2.2.2.1 Overview

System level and DAAC management and technical representatives developed the V0 metadata model, a *de facto* metadata content model used in the Information Management System (IMS).

http://www-v0ims.gsfc.nasa.gov/v0ims/DOCUMENTATION/GUIDE-VALID/VALIDS/validauthors.html

2.2.2.2 Process and Evolution

Management and technical representatives from all eight Distributed Active Archive Centers (DAACs), with several system-level team members, built a consensus on common search attributes in a few initial meetings. Subsequent discussions occurred at semi-monthly teleconferences and on developer email lists. Development and science or user services teams, DAACs, and system level teams reviewed the set of controlled keywords periodically. Developers consulted with DAAC scientists as needed, particularly on valid search values (“valids”). The initial metadata model took about 6-12 months to develop.

Community Agreement: Consensus from DAAC management and the developer community was reached at the initial development meeting and through subsequent teleconferences.

Eventually, the IMS development team agreed to rely on the GCMD keyword lists where applicable, greatly simplifying keyword maintenance and cleanup of "valids."

2.2.2.3 Tools and Documentation

A link to "valids" information is available on the Web at http://www-v0ims.gsfc.nasa.gov/v0ims/DOCUMENTATION/documentation.html

2.2.2.4 Current Status

The V0 metadata model has nine basic search attributes, just enough to perform system-wide search and order activities through a common Information Management System (IMS). Additional attributes are available for data product descriptions, ordering options, and specialized, dataset-specific search criteria. DAACs map the system metadata model to their own individual, heterogeneous metadata models.

IMS is required as a metadata interchange format within the V0 IMS system, which also serves as the ECS user interface. As such, it is used by all NASA DAACs and various other IMS nodes internationally. The IMS community has moved closer to the GCMD model and keywords and is working with the CEOS community on harmonization.

2.2.3 ECS Metadata Model

2.2.3.1 Overview

NASA's Earth Observing System Data and Information System (EOSDIS) Core System (ECS) was designed to ingest data volumes on the order of 2TB per day. Incoming data granules are categorized by time, space, altitude, data type, etc., and undergo calibration and re-processing within the archive. This requires a certain set of metadata; another set is needed for searches and retrievals by scientists from diverse disciplines, and for access by none-experts such as lawyers, policy makers, educators, etc. Producers and inter-disciplinary data users have quite different requirements for searching and ordering; users may request a single granule, a collection of similar granules; browse or descriptive information instead of (or prior to) the data; or information on production history, storage format, and production algorithms.

The ECS metadata design sought to meet this broad set of requirements. The expected diversity of uses, and the volume and number of granules, led EOSDIS to design a complex data model that describes data holdings and organizes them into collections of similar granules, represented by Earth Science Data Types (ESDTs), each with attributes describing the data's origin, its spatial and temporal coverage, contact information, and other descriptive elements. Now also an FGDC standard, the ECS metadata content model defines over 300 attributes (plus many "Product-Specific Attributes") with a "minimal set" of about 15 required for all data sets.

http://ulabibm.gsfc.nasa.gov/metadata/newmdata/index.html

2.2.3.2 Process and Evolution

The EOSDIS team conducted an extensive survey of existing metadata models and extensive interactions with science community in an attempt to capture and describe all possible data attributes. The initial metadata model took about three years to develop. The team inserted additional features, such as geographic description paradigms, as required by science teams.

2.2.3.3 Tools and Documentation

Data providers have two tools for help in generating metadata that conforms to the ECS metadata model. The Science Computing Facility Toolkit, a subset of the Science Production Toolkit, provides a library of functions that can perform various tasks, including Creation of properly formatted metadata files. Meta-DataWorks is a World Wide Web-based ESDT Collection / Granule Metadata Population System used to create ESDT descriptor files and the associated Metadata Configuration Files (MCFs) required by the Science Computing Facility (SCF) Toolkit for processing.

2.2.3.4 Current Status

The requirement was for a single metadata model to cover all the datasets to be archived in ECS. A comprehensive ECS Metadata standard was developed using extensive interviews with the science community and complying with the FGDC content metadata standard. This substantial job took a long time and produced a standard that is complex, but has a minimal number of required fields. This metadata model is required for data archived in ECS.

ECS metadata actually is a good metadata standard for remote sensing where there is a history of needing complex but comprehensive metadata.

However, there is a general impression that it is too complex to be used. There are several reasons for ECS metadata not being widely used:

- There are no easy-to-use stand-alone tools available. The tools from ECS were embedded in large data system not intended for use outside of ECS.
- The promotion of the standard was not very effective in the past. This situation is changing.

The standard, with some revision and extension, is becoming an FGDC standard and will eventually become an ISO standard. OGC is starting to use it in its IP program. Therefore, we may well see more use of the standard in the future.

2.2.4 Lessons Learned – Metadata Standards

The *simplicity* of the metadata model facilitates building consensus on common data attributes, while the *flexibility* of the model provides support for dataset-specific attributes as necessary.

2.3 Catalog Interoperability Standards

2.3.1 V0 Information Management System (IMS)

2.3.1.1 Overview

The V0 IMS Protocol is a simple, stateless, message-based protocol. The communications protocol and message suite are defined using the Object Description Language (ODL), a parameter-value language originally developed to describe the form and content of scientific data in the NASA Planetary Data System. The V0 IMS adopted the freeware ODL Processing Software Library for creating and parsing IMS messages. The V0 IMS provides search and order capabilities and has been successfully used as a cross-DAAC search and order layer and by international data centers interoperable with V0. The V0 system has been operational since August 1994.

http://redhook.gsfc.nasa.gov/~imswww/pub/imswelcome/

2.3.1.2 Process and Evolution

Developers from all NASA DAACs and the V0 system level team (all funded from the same source) met twice a year for technical discussions of needed capabilities and possible solutions. At these meetings, they made decisions and implemented them immediately. The V0 systems level team played a leading role in the meetings by guiding technical discussions. Subsequent discussions occurred at the semimonthly teleconferences and on the developer email lists. Science community involvement occurred through four “tire-kickers” who attended the biannual development team meetings to give science input on the new capabilities.

Initial protocol decisions were based on technical requirements and availability of software tools for all DAAC platforms. Extended or modified capabilities of the V0 IMS protocol were handled at the semimonthly teleconferences and the biannual development team meetings. Continued protocol evolution falls to a “tiger team” consisting of V0 system-level team members and the group requiring the extension, and reviewed by the development team. The initial V0 IMS protocol took a small technical group about six to twelve months to develop.

2.3.1.3 Tools and Documentation

IMS messages are generated and parsed using the ODL Processing Software Library, originally developed for NASA’s Planetary Data System. Thus, while the PDS project standard for data description was not re-used for this project, the supporting software library was. The V0 system level team designed, implemented, and tested the software library that performed the V0 IMS communications functions. The V0 System level team distributed the both ODL and communications software to the DAAC developers, who were responsible for integrating the software with their V0 Servers. The V0 system level team maintains the “IMS Messages and Development Data Dictionary”. The V0 Server Cookbook provides basic information on “plugging in” a new data server, with links to tools and sample software.

http://www-v0ims.gsfc.nasa.gov/v0ims/DOCUMENTATION/DDICT/edg_ddict.html

http://www.esipfed.net/v0/server-cookbook.html

2.3.1.4 Current status

The availability of a clear interface, documentation, and reusable software Server and Client components, as well as the simplicity of the interface and strong project leadership, has led to adoption of the V0 interface by the EOSDIS DAACs and six to eight independent international partners.

2.3.2 V0 Guide

2.3.2.1 Overview

The V0 IMS Guide Protocol is based on the library community’s catalog search and retrieval protocol, Z39.50, as well as HTTP, to provide search of detailed data description documents, using free text and fielded keyword search.

2.3.2.2 Process and Evolution

The GSFC DAAC played a leading role in identifying the Z39.50 and HTTP standards as the standards to adopt for Guide search and access, and led the periodic Guide development meetings by steering technical discussions. Selection of the standard protocols took 2-3 months.

The Guide system has been modified with additional capabilities and updated with more recent versions of the Z39.50 protocol and software components. The process described above was used for the evolution of the Guide system.

2.3.2.3 Tools and Documentation

A lead DAAC used existing Z39.50 software components and then modified them for use as a Guide server. Adopting an existing international standard that had freeware software components available accelerated the development cycle for the Guide Server. The completed Guide server component was distributed for installation at multiple DAAC sites and clear written documentation was provided. A variety of guide document authoring tools were developed and distributed to the DAACs to use to populate the Servers with Guide documents. However, these were neither well supported nor widely used. Instead, well documented

Guide document templates and search code lists have allowed DAACs to develop their own document handling scripts.

http://www-v0ims.gsfc.nasa.gov/v0ims/DOCUMENTATION/GUIDE-VALID/GUIDES/guideauthors.html

2.3.2.4 Current Status

All EOSDIS DAACs, and several international data centers, have adopted the V0 IMS Guide. have adopted it as well. The use of the Z39.50 international standard has enabled the Guide system to be linked to the Federation Geographic Data Committee (FGDC) Clearinghouse and the implementation of the National Spatial Data Infrastructure (NSDI), which also is Z39.50 based. This is expected to generate more user access to the Guide documents.

2.3.3 Federation Interactive Network for Discovery (FIND)

2.3.3.1 Overview

The Federation of Earth Science Information Partners (ESIPs) adopted two complementary search systems for collection level searching: the GCMD, NASA's comprehensive directory system for global change data, and Mercury, a less-structured search system for Earth science data which is based in Internet search engine technology, developed and maintained by the ORNL DAAC. The EOSDIS V0 IMS was adopted for inventory level search and order.

http://esipfed.net/committees/interop/swil.html

2.3.3.2 Process and Evolution

The Federation Interoperability Group met via monthly teleconferences and briefly at the twice-annual Federation meetings. A small "tiger team" drafted evaluation criteria for defining a Federation System Wide Interface Layer (SWIL) – now known as the Federation Interactive Network for Discovery (FIND). A two-day workshop was held in conjunction with a Federation meeting and the draft SWIL Evaluation Criteria was reviewed and consensus agreement was reached. An RFP was issued for the Federation SWIL implementation. Multiple proposals were received and evaluated, and recommendations were passed to the Federation Appropriations Committee, who passed them on to the NASA Program Manager. The NASA Program Manager performed an external review of the proposals, reviewed the Federation Interoperability Team recommendations, and then funded several proposals for implementation. The entire process of selecting SWIL technology was lengthy and took about two years. In the end, previously existing, already implemented technology was adopted and minimal new capabilities were developed The Federation Interoperability chair and a few key people provided the strong system level project management support needed to push the process through.

Evolution is difficult. The extremely lengthy process poses a high barrier to adding small extensions or modifications in capabilities to the existing systems.

2.3.3.3 Tools and Documentation

GCMD and V0 IMS provide tools and documentation as described above. Mercury harvests metadata from the GCMD and from ESIP Web sites (no tools are required for an ESIP to participate at this level), and provides metadata entry tools and accompanying documentation for those ESIPs wishing to provide additional metadata. “Plug in” information for all three systems is provided from the Federation web site: http://esipfed.net/committees/interop/swil.html.

2.3.3.4 Current Status

All the Federation ESIPs are required to define their data collections through the GCMD DIFs and do. This was an original requirement in the cooperative agreements that funded the ESIPs. About half have also implemented the V0 IMS interface or provided their metadata through a DAAC for inventory searching.

2.3.4 Lessons Learned – Catalog Interoperability Standards

Widespread adoption is more likely if reusable software components are available and the interface is simple, well documented and easy to understand. Continued technical support, both at the system level for system maintenance and evolution, and for new adopters of the interface, is also important. Note: After V0 finished development, many scientists outside the four “tire-kickers” made unfavorable comments on the need for a V0 inventory system. It’s difficult to get consensus on information systems by the science community.

It is also important to keep abreast of new technology. Very often, an independently developed *de facto* “standard” will emerge just when it is needed[2]. Note that at the time they were selected for the V0 IMS Guide, HTTP and Z39.50 were very new protocols. It wasn't until we saw Marc Andreesen (then a college student) demo Mosaic at a workshop that we made a decision, and it was not a completely obvious choice at the time. We could easily have guessed wrong on HTTP…

Strong system level project management support is critical to adopting system level interfaces. Building consensus among many disparate, competing organizations is difficult. Community agreement doesn’t guarantee adherence to the agreements. A long RFP process for selecting new capabilities is not evolution-friendly for small modifications and extensions.

2.4 Data Access Interoperability Standards

Data access interoperability will allow software applications to access diverse data at multiple sites using a common interface. A variety of data access

[2] While this is based on anecdotal evidence, remember that necessity is the mother of invention and often different groups or projects have similar needs and are working on solutions to similar problems.

interoperability tools are currently available, which may or may not support open standards and interfaces. Several of these technologies are in use within the ESIP Federation, as described in the "Interoperable Data Services for Earth Science Data" white paper: http://esipfed.net/committees/interop/swil.html.

In exploring system-wide interoperability issues, the ESIP Federation found it much easier to reach consensus on high-level catalog interoperability (described above) than on data access. This was due in part to the relative maturity of technology supporting these two types of interoperability, and in part to the diversity of the ESIPs and their user communities. Instead of trying to mandate a single interoperable data access protocol, the Federation formed "clusters" of ESIPs with common interests, such as interoperable data access technologies. Within the Federation, the clusters provide a testbed for exploration of new technologies, and a source of information and technical assistance for the spread of these technologies throughout the Federation. An example system, which is standards-based, is described below as an example of a standards adoption process.

2.4.1 Federation Geospatial Services Cluster

2.4.1.1 Overview

The Geospatial Service Cluster is a group of ESIPs who are exploring Web Mapping technologies for interoperable data access in order to provide data to the GIS and education user communities. The cluster is adopting OGC web mapping technologies as the standard for interoperability. Initial work within the cluster has focused on the Web Map Server specification, which provides visualization of the data (maps). As additional OGC web mapping standards are finalized, such as Web Coverage Server to provide access to actual data values, this cluster will further its prototyping efforts.

http://oceanesip.jpl.nasa.gov/esipde.html

2.4.1.2 Process and Evolution

The cluster of interested ESIPs formed in summer 2000 around a mutual interest in the Digital Earth program. Rather than developing an interoperable data access standard, the cluster is prototyping Web Mapping standards as they evolve within OGC.

2.4.1.3 Tools and Documentation

2.4.1.4 Current Status

There has been wide interest in this technology within the Federation, and most of the ESIPs participate to some degree in this cluster.

2.4.2 Distributed Ocean Data Service (DODS)

DODS is another popular data interoperability system, not originally developed based on open standards, but moving toward "standardizing" their Data Access Protocol.

2.4.3 Lessons Learned- Data Access Interoperability Standards

The availability of tools and investment by the wider NASA ES community and commercial sectors have been key to this success.

2.5 Data Format Standards

2.5.1 HDF and HDF-EOS

2.5.1.1 Overview

HDF is a file format and software library for sharing data in a distributed environment and is especially oriented toward storing scientific data. It is a self-describing, well documented, and portable data format, which is well suited to sharing data within heterogeneous computing environments. HDF is developed at (and "owned by") NCSA. HDF-EOS, an extension of HDF, is developed and maintained by ECS in conjunction with NCSA.

http://hdf.ncsa.uiuc.edu/

http://hdfeos.gsfc.nasa.gov

2.5.1.2 Process and Evolution

In the early stages of the EOSDIS development, NASA defined a process for selecting a standard data format. The process included prototyping several Earth science data formats with the help of DAACs and other user communities. A document was published describing the whole process and HDF was selected as standard for the EOSDIS project. The process was based on consensus and prototyping. This data format standard selection process took about 6-12 months.

Development of both HDF format and supporting tools in response to science community needs has continued and has spawned the HDF-EOS format, maintained by an ECS contractor. Built "on top of" HDF, HDF-EOS supports grid, swath, and point objects (composed of standard HDF data structures), along with structural metadata. A new version of HDF-5 and companion HE5 are in the works. HDF-5 shares the same data models with HDF-4, e.g., it implemented the swath, grid, and point data models of EOSDIS. However it is incompatible at the source code level in terms of implementation of data structures, user interface, and even the data storage paradigm. HE5, however, is supposed to be source-level compatible with HDF-EOS 4.

2.5.1.3 Tools and Documentation

A variety of tools and complete documentation are available from NCSA. In particular, tools for translation between HDF-4 and HDF-5, as well as several papers, which address the changes required to migrate software and tools from HDF4 to HDF5, are available on the HDF Web site. The HDF-EOS Web site provides links to ECS-developed and third party tools as well.

HDF is also built into a number of popular COTS tools (IDL, MATLAB), and is supported by (interoperable with) other standards such as DODS.

2.5.1.4 Current Status

HDF and HDF-EOS are widely used for NASA mission data. However, user communities may request data to be translated to simpler binary or ASCII formats before distribution. HDF/HDF-EOS acceptance has evolved as Landsat 7 and Terra data are distributed and more tools are available. A user survey by the National Snow and Ice Data Center (NSIDC) found that a significant majority accepted HDF, and half preferred HDF over other formats.

2.5.2 Lessons Learned- Data Format Standards

Many science communities have collective experience and tools for other formats (e.g., netCDF, GRIB). Costs for conversion of legacy data are high. Because a single data format is not always practical, this is a good candidate for a community standard, rather than a core standard.

2.6 Data Exchange Standards

Exchange of data among data centers, between data producers and data centers, and between data centers and their users is accomplished by a variety of tools and protocols, many informal or developed specifically for the job at hand. These individual data exchange methods have varying degrees of automation and reliability. Systems that exchange data with a variety of others have an interest in standardizing the process in order to reuse data exchange, ingest, and distribution software to the maximum extent possible.

2.6.1 ECS Interface for External Providers

2.6.1.1 Overview

This interface was designed to be used by external data providers to provide their data through ECS and for external client developers to develop alternate user interface clients to ECS. This interface was overly complex.

2.6.1.2 Process and Evolution

From requirements, the ECS contractor designed and developed the interface and then provided information about it to the project and external organizations. There was no community agreement.

2.6.1.3 Tools and Documentation

2.6.1.4 Current Status

No one has used this interface. This interface was abandoned after a period of time.

2.6.2 ECS Interface for Science Investigator-led Processing Systems (SIPS)

2.6.2.1 Overview

This interface moves data from a science-investigator-led processing system to a Distributed Active Archive Center.

2.6.2.2 Process and Evolution

The interface came into being when processing was outsourced to PIs, and was based on DAN/DDN (TSDIS-GDAAC), which was based on SDPF CBI protocol. It replaces the ECS Interface for External Providers in functionality.

2.6.2.3 Tools and Documentation

No tools are available. The interface protocol was fully documented in SIPS/ECS Interface Control Document.

2.6.2.4 Current Status

This fairly simple, structured, and reciprocal interface, designed specifically for this job, works well. The various SIPS use it in sending data to DAACs.

2.6.3 ECS Machine-to-Machine Gateway

2.6.3.1 Overview

The interface allows SIPS to search and order data under programmatic control, primarily for reprocessing campaigns and quality assurance.

2.6.3.2 Process and Evolution

It became necessary when the SIPS-DAAC interface came into being. Implementation descends from defunct JESST interface.

2.6.3.3 Tools and Documentation

A document describing the interface is available from ECS.

2.6.3.4 Current Status

The emphasis on reuse of existing ECS functionality makes this interface somewhat awkward. Request, response, and data delivery functions use two different communications mechanisms (ssh and email), with three different “language” syntaxes (MOJO, XML or ODL, and the format defined for data delivery emails). Some systems (e.g., external subsetters) have elected to order data using the V0 protocol instead.

2.6.4 Lessons Learned – Data Exchange Standards

Ideally, the process needs community agreement during development and the method needs simple interfaces with developed tools and software components.

Interfacing with a complex entity (large organization, large established code base, etc.) is difficult, especially if such external interfaces were not planned originally. The complex entity will want to (and may need to) dictate the interface standards.

Reuse of existing interfaces may be the best way to extend a complex system like ECS. For example, the integration path to provide for sub-setting and other processing external to ECS-supplied software involves reuse of a variety of existing interfaces to ECS in order to lessen system development costs. These interfaces include the V0 IMS protocol for specification of custom processing requests, requesting granules to process from ECS, and providing processing status updates to ECS; ECS data delivery interfaces (FTP push and Delivery Notification) for acquiring data from ECS for processing; and a modified SIPS interface for returning custom-processed data to ECS for subsequent delivery to the user.

Nevertheless, if the external interface is designed entirely for the convenience of the complex organization, it may not be practical for the other party, and so it won't be used.

SEEDS will incorporate organizations of different sizes and with different heritages, so we may run into many variations on this situation. By extension, if SEEDS needs to incorporate several "800 pound gorillas" that want external interfaces on their own terms, this will lead to several competing "standard" ways to provide these interfaces....

2.7 NASA-ESE Standards in Progress

2.7.1 ECHO

2.7.1.1 Overview

The EOS ClearingHOuse (ECHO) (http://dangermouse.gst.com/echo/) is an enabling framework that allows for interoperability between data, services, and clients via application program interfaces (APIs).

ECHO presents an open API based on XML. The ECHO model is based on e-commerce technologies and is reusable for many applications (Earth science and other). For the current Earth science application, it is a clearinghouse of metadata that describes data at both a collection and granule level. Metadata also describe services, clients that can operate on the data and services, and the providers who participate. ECHO uses an open system approach. ECHO ensures that user interfaces fully address user needs by specifying and publishing domain APIs that accommodate independent clients. APIs are in-dependent of the underlying transport protocols used. They are capable of communicating using SOAP and Java RMI. Other transport protocols can be added as necessary.

2.7.1.2 Process and Evolution

The ECHO advisory group provided the development team with a set of use cases describing the system. The development team then produced a set of APIs and presented them. The team does its development in iterative fashion, allowing community feedback from workshops, telecons, and regular meetings of the ECHO Technical Committee (ETC) to guide what parts get built next.

The process for adopting/establishing ECHO standards is through a technical study team (ECHO Technical Committee / ETC) in conjunction with the develop-ment team. The ETC includes representatives from each provider, the develop-ment team, and ESDIS. Some ETC members also serve as interfaces to key external standards groups (e.g. OGC) to ensure interoperability in a broader context. The development contractor or a member of the study team presents standards options to the ETC along with a recommended standard. Discussion takes place and then the standard is voted on. ECHO is developed in increments to allow for insight and feedback during the development cycle. Industry trends are exploited when opportunities present themselves. This optimizes the use of COTS products.

2.7.1.3 Current Status

The metadata standard is being evolved from the ECS data model. APIs are based on a simple XML message-passing interface. Data Type Definitions (DTDs) define what messages look like. Configuration files route XML messages to the appropriate business logic, allowing the simple addition or modification of messages. Services encapsulate groups of related transactions and control the access based on the roles of connecting entities. ECHO supports or envisions several types of services, including internal services (e.g. services on or in the clearinghouse), data services, administrative services, and search services.

ECHO has leveraged a number of standards in producing its system. XML is the basis for all communication with ECHO. At the project's outset, XML had just become a standard, and there were no extensions of it standardized. ECHO adopted XML as the communication mechanism for the system.

ECHO adopted the OGC proposed XML catalog service structure with some modifications. This has formed the basis for interacting with the catalog of meta-data stored in ECHO. ECHO also adopted GML as one method of describing spatial data in the system. ECHO also leverages the Oracle spatial capabilities, which are also related to OGC activities.

2.8 Recommendations

Key factors observed for widespread adoption of a standard or interface within the NASA EOSDIS:

- Community involvement during the development process
- A small group of people involved during development process
- Strong project management staff needed to lead the technical discussions, the implementation, and overall management essential
- Software tools and components readily available. This may require NASA investment for user/research communities to develop tools and/or provide technical support.
- Simple interfaces

Overall recommendations:

- The standard needs to be simple. The more complex or large the standard and more difficult it is to understand, implement, and maintain, the less likely the widespread adoption of the standard.
- An extensive set of standards-based tools and software components is a necessity for acceptance.
- The motivation to work together is important. This motivation can result from having common funding or it can come from having common objectives.
- Adopting a standard where software components are available can shorten the development and implementation cycle.
- A spiral development process is a more effective process than the traditional waterfall method.
- Multiple interface standards are needed to address the requirements of individual user communities. SEEDS will serve many different user communities, such as Ocean, Atmosphere, Land, etc. Each community has different de-facto communities standards. It is essential for making SEEDS successful

that SEEDS should have multiple interfaces to address the community standards although SEEDS can have a set of core standards that may be different from any community standards.

- Community agreement doesn't guarantee community adherence to the agreements.
- For each standard, developers need to identify who the people are that represent the community. Is it DAAC developers? Is it instrument scientists? Is it science PIs? Is it users? Standards that may be extremely advantageous to one group may be extremely burdensome to another group.
- It's difficult to get consistent input/feedback from the science community over a sustained period of time on information system interfaces, even from the same individual scientists.
- Negotiation of system standards is easier with a smaller community and becomes more difficult as the size of the community increases.
- Negotiation of system standards is easier with a small, homogenous community that uses common terminology and has a common base of knowledge. Negotiating with a large group of people from different backgrounds (technical, different science disciplines, etc.) on information system interfaces is almost impossible.

3 OTHER EXPERIENCES WITH STANDARDS & INTERFACES

This chapter offers some points of comparison with NASA's own experiences in terms of creating and deploying information standards and interfaces. It reviews the experiences of Canada's GeoConnections program, as well as the US National Oceanographic and Atmospheric Administration (NOAA), the global Grid Computing initiative, and Sun's Java Community Process.

3.1 Canada's GeoConnections Program

3.1.1 Overview and Description

The Canadian government initiated the GeoConnections program in 1998 to facilitate and enable Internet access by Canadian users to all types of geospatial data and information produced by Canadian federal, provincial, territorial governments, and to encourage new uses of this data and information. The GeoConnections Program will build the Canadian Geospatial Data Infrastructure (CGDI). It supports five policy themes:

- Ensure 24/7 geospatial data access over the Internet by all Canadian users.
- Establish a framework of data to facilitate information integration for rapid decision making and development of new information products.

- Use international geospatial information system standards to facilitate sharing information with other nations and to ensure that Canadian businesses can sell geospatial information technology and services in the global marketplace.
- Collaborate in partnerships with all levels of government, industry, and academia to leverage their expertise.
- Develop a supportive policy at all levels of government to accelerate private sector commercialization of geospatial data and information.

GeoConnections has seven funded programs to develop the Canadian Geospatial Data Infrastructure (CGDI), which support the five policy themes.

- The *Access Program* is a partnership with all levels of federal, provincial, and territorial governments to make geospatial data Internet accessible.
- The *Framework Data* program is constructing a geospatial data framework to enable application and value added development. The framework will be formed with diverse data from multiple organizations upon which other databases will be integrated or constructed. The Framework program involves complex negotiations with the provinces and is directly related to the Access Program and evolving policy issues. (E.g. framework dataset of roads in Canada)
- The *GeoPartners* program coordinates federal, provincial, and territorial ideas, programs, and project activities to guide the vertical harmonization of information sharing and access to geospatial data and information.
- The *GeoInnovations* program funds promising technologies and tool development by the industry
- The *Sustainable Communities* program builds or strengthens the capacity of Canadian communities to plan and manage their economic, environmental, and social development using geospatial data, information and services offered on the Internet.
- The *National Atlas of Canada* program uses the Internet to provide perspectives on Canada's physical, environmental, economic, social, and cultural issues for students at all levels and the public in general. This program provides online views of geospatial data and information and places provincial and territorial data in a national context, illustrating the regional dimensions to the issues. (e.g. climate change, population density)
- The *Geomatics Skills Network* is a skills matching program to sustain industry growth and fill job vacancies.

Several other unfounded programs complement the initiative by providing strong policy and technical guidance or application nodes that involve user communities such as transportation, sustainable development, disaster management and marine biology.

Of the seven funded programs, the Access Program and the GeoInnovations program are most relevant to this study. The Access Program will be discussed in more detail in the following subsections.

3.1.2 Problem Statement: The GeoConnections Access Program

The GeoConnections Access Program (Node) is a partnership with all levels of federal, provincial, and territorial government to make all government geospatial data accessible via the Internet, thereby developing portions of the Canadian Geospatial Data Infrastructure (CGDI). The Access Program is much broader than a traditional clearinghouse: it provides capabilities for users to discover, evaluate, and access geospatial data and information using multiple Internet based tools and services, and mechanisms for data providers to advertise and distribute their data and services.

3.1.3 Constraints of the GeoConnections Access Program

At the initiation of the GeoConnections program, the Treasury Board (similar to the US General Accounting Office) and the GeoConnections Management Board defined a set of program constraints for the Access Program as it consults, develops, integrates, operates and forms partnerships:

- The Access Node Advisory Group represents the interests of industry, academia, federal and provincial government, and user groups.
- A Technical Advisory Panel (TAP – an unfunded node), composed of technical experts from industry, academia, and government, gives advice on the technical strategic direction and evolution of the GeoConnections program for the development of the Canadian Geospatial Data Infrastructure.
- It was decided early on to adopt only international standards (formal or *de facto*) for the GeoConnections program. This would help the Canadian industry become players in the global marketplace and also allow access to the Canadian data and information by international users. Standardization also enables interoperability with other countries, with a view to building a Global Spatial Data Infrastructure.
- GeoConnections funding was allocated at the beginning of the program for the full five year program period. A decision was made that a competitive RFP process would award funding to partner data providers who could meet the terms of the RFP to put spatial data online.
- The GeoConnections Program is national in scope, hence respects the visibility of other federal departments and provinces or territories. Therefore, CGDI will be a distributed system of systems.
- GeoConnections will add desired capabilities into already existing, heritage systems whenever available.
- GeoConnections followed the incremental development style, using multiple contract vehicles.

3.1.4 Process

About two dozen staff work with four contractors on development in the Access Program. Another 7 people work on operations and connectivity issues, e.g. metadata entry/quality control, connecting distributed servers, user support, etc. -- for a total staff count of less than two dozen. In addition, the Access Program has a partnership program that receives contributions and resources from other federal and provincial government organizations. The “central” team mentioned above coordinates this broader activity.

A Technical Advisory Panel (TAP), composed of about 30 technical experts from industry, academia, and different levels of the Canadian government, chaired by a government representative, advises all 7 GeoConnections programs. The TAP initiated and managed the Target Vision and Architecture work and makes recommendations for standards to be adopted. The Access Node Advisory group usually meets with TAP to facilitate a free exchange of information and advice. TAP members meet twice a year to review the program and make technical decisions. Decisions are by consensus. TAP has working groups that work between meetings.

Initially, a subgroup of the Technical Advisory Panel met to define a target vision and conceptual architecture. It funded three competitor companies to develop independent conceptual architectures, and combined the three studies into a single conceptual architecture. The entire process took 1 to 1.5 years. Currently the CGDI architecture working group conducts ongoing activities related to CGDI architecture and service definition. Detailed information on this group can be found at: http://www.geoconnections.org/architecture

There have been a number of reviews of the target vision and architecture; an initial “public review” of the work of the 3 contractors, and reviews at the Node Advisory meetings. After the conceptual architecture was defined, TAP developed an implementation plan.

TAP recommended adoption of 5 interface standard specifications from the OGC and ISO standards organizations for use in CGDI: OGC Catalog Interface, OGC WWW Map Server, OGC WWW Feature Server, OGC Geography Markup Language, and the ISO TC 211 19115 Metadata Standard. The other GeoConnection programs concurred, and accepted these interface standards as GeoConnection standards.

The Access Program staff issued a competitive “Call for Participation,” targeted at federal and provincial government organizations, in October 2000 with an end date of March 2005, to provide up to 50% funding to connect data providers to the CGDI using the interface standards. All GeoConnections funding requires the recipient / partner to provide at least 50% of the cost share. A negotiated process results in an agreement between the Access Node and the other government organization. A successful agreement requires that the bulk of the funding go to industry partners, but it is the responsibility of the federal or provincial partner to contract with industry.

Community engagement for data providers occurred through a variety of means: RFPs, direct funding transfers to government suppliers, use of government staff to help connect suppliers, and standing contracts for connectivity support on an "as and when" basis. Perhaps the most effective engagement is at the database administration level where each metadata coordinator is in daily contact with many data providers to encourage them to add data collection descriptions to the directory, to update existing entries, and to connect distributed inventories. Many data providers create high level data collection metadata descriptions in the GeoConnections directory (similar to the GCMD) with no funding from the program and a few suppliers have connected distributed databases with no funding from the Access Program. Data providers have been supportive of the Access Program because they see clear benefits to themselves and the buy-in cost is very small.

3.1.5 Standards Relevant to ES

GeoConnections Access program has identified 5 standards and interfaces for adoption across the CGDI:

- OGC Catalog Interface
- OGC WWW Map Server
- OGC WWW Feature Server
- OGC Geography Markup Language
- ISO TC 211 Metadata standard

3.1.6 Lessons Learned

GeoConnections lessons:

- GeoConnections was able to obtain the entire 60 million dollar, multi year funding all at once, and could thus build a broad program and work on many related tasks in parallel.
- GeoConnections never funded more than 50% of the activity. Partners must obtain funding for the other 50% from other sources. This filters out less committed partners.
- Using funding to ensure compliance with the interoperability standards was key to the standards compliance by data providers and industry partners

Access Program lessons:

- Using an already existing information system, CEONet, was very helpful to jumpstart the CGDI implementation.
- Incremental development was essential in developing the CGDI.
- Loose coupling between data providers, with integration only through common interfaces, made it easier to engage data providers and cheaper to manage and administer the infrastructure.

- Access node development leveraged the GeoConnections staff international experience, primarily with CEOS. The Access Program still benefits operationally from participation in the IDN where monthly updates of international directory entries are received.
- Adopting an accepted standard like the FGDC/GEO enabled the Access Program to jumpstart itself. Initially the Isite software from FGDC was used for immediate interoperability with a large number of distributed catalog servers in the US and internationally.

3.2 The US National Oceanographic and Atmospheric Administration (NOAA) server

3.2.1 Overview

NOAA server (http://www.esdim.noaa.gov/noaaserver-bin/NOAAServer?stype=-home) is an operational system built on Web based standards with the goal of providing an integrated view of many data systems across NOAA. The process of evolution of NOAA server is internal, with participation of many participants from NOAA centers. NOAA server came into existence as a result of an action item from NOAA administrator as part of GCRP program to address IT issues and standards. The initial development started in the 1990. This was part of a tri agency (NASA, NOAA and USGS) GCDIS plan. The planning process and evolution of the system was influenced by budget constraints. NOAA server attempted to provide a common look and feel for many diverse systems relying on web based standards and protocols. Currently there are 14-16 nodes similar to FGDC clearinghouse nodes. The NOAA server development process was influenced by budget, diversity of mission, corporate culture and multiple organizations with different requirements.

3.2.2 Process

NOAA server is part of an overall NOAA infrastructure comprised of many small autonomous systems. Because of the nature of organization, the system development and adoption of any process needs to take into consideration organization and territorial issues.

The development process was through periodic meetings and was an evolutionary process. There was very little user community involvement and focus was mostly on internal users. The NOAA server had a senior advisory community which made decisions based on consensus and guided the use of standards and technologies. This was primarily an integration effort leveraging existing systems. It was decided to use FGDC metadata standard and Web based standards. Search engine was based on WAIS and Z39.50. The team also investigated CORBA. Currently NOAA server management is investigating OGC complaint standards implementation and also investigating Arc IMS and Arc View. There were no specific metrics to measure the successes of the system except informal user statistics.

3.2.3 Lessons Learned

- Cultural differences across multiple organizations
- Competing approaches
- Success depends on people and organizations contributing data to the system
- Limits of consensus seeking process – No clear path to substitute this.
- Struggle between technology evolution and operation
- Incremental approach works best
- Difficult to define technology infusion for ongoing process

3.3 Global Grid Computing

3.3.1 Overview and description

The continuing decentralization and distribution of software, hardware, and human resources make it essential to develop new approaches in computing. This requires new concepts and strategies that let applications access and share resources and services across distributed, wide area networks, while providing common security semantics, distributed resource management, and performance. Work within the framework of Grid technologies for scientific research addresses issues related to Grid computing.

The Global Grid Forum (GGF) has become the central organization to coordinate the efforts of developing various aspects of "Grid" computing and has grown out of a series of conversations, workshops, and Birds-of-a-Feather **(**BOFs). The GGF is a community-initiated forum of individual researchers and practitioners working on distributed computing and data, or "grid" technologies. Efforts include those specifically aimed at very large data sets, high performance computing but increasingly those efforts that industry is calling "Peer-to-Peer. GGF is the result of a merger of the *Grid Forum*, the *eGrid European Grid Forum*, and the Grid community in Asia-Pacific. GGF participants come from over 200 organizations in over 30 countries, with financial and in-kind support coming from GGF Sponsor Members including commercial vendors and user organizations as well as academic and federal research institutions**.** The GGF mission is "*to focus on the promotion and development of Grid technologies and applications via the development and documentation of "best practices," implementation guidelines, and standards with an emphasis on "rough consensus and running code".* The GGF is playing a major role in the co-ordination and development of standards, interface and infrastructure for the Grid computing technologies (http://www.gridforum.org).

Some of the major funded activities in the US include NSF funded Tera Grid-TeraGrid is a multi-year effort to build and deploy the world's largest, fastest, most comprehensive, distributed infrastructure for open scientific research. When completed, it will include 13.6 teraflops of Linux Cluster computing power distributed at the four major TeraGrid (http://www.teragrid.org) sites: the National Center for Supercomputing Applications (NCSA), the San Diego Supercomputer Center (SDSC), Argonne National Laboratory, and the California Institute of Technology (Caltech). Other examples are NASA's Information Power Grid (http://www.ipg.nasa.gov/) , Grid Physics Network (www.griphyn.org), and Network for Earthquake Engineering Simulation NEESgrid (www.neesgrid.org), and DOE science Grid (http://doesciencegrid.org/).

The purpose of this article is to primarily focus on the process involved within GGF which is achieved through consensus and community participation towards the development of Grid technology , standards and policies. We will not discuss all the Grid activities here. Those who are interested to find a detailed description of all the Grid activities should visit http://www.gridcomputing.com/.

GGF Structure

GGF is managed by a Steering Group (SG) led by a chairman. An external advisory committee (AC) provides long-term strategic input to the SG.

GGF working groups and research groups are organized into "areas" of interests each managed by two members of the SG (Area Directors). To manage the logistics of GGF's activities, GGF established a fully incorporated not for profit business entity, **GGF, Inc.**, in July 2001.

GGF structure is documented in "Structure: Areas, Working Groups, Research Groups."

GGF Governance is documented in http://www.gridforum.org/Documents/GWD-C/old/GGF-Management-22-Oct.pdf

A document describing the GGF management structure and process in detail is available (http://www.globalgridforum.org/L_About/Struc_Proc.htm).

3.3.2 GGF Process

GGF meets 3 times per year at different locations. GGF is in the process of creating a persistent document series for grid computing that is analogous to the Request for Comments (RFC) series associated with the Internet Standards Process and the Internet Engineering Task Force (IETF). A draft process by which this document series is managed has been developed by GGF leaders, loosely modeled after the Internet Standards Process as described in RFC 2026, "The Internet Standards Process- Revision 3" by Scott Bradner (October 1996).

Further information on the procedures followed by the Working groups can found in the following references:

Bradner, S., "The Internet Standards Process- Revision 3," RFC 2026, October 1996

Weinrib, A. and Postel, J., "IRTF Research Group Guidelines and Procedures", RFC 2014, October 1996.

Bradner, S., "IETF Working Group Guidelines and Procedures", RFC 2418, September 1998.

Postel, J. and Reynolds, J., "Instructions to RFC Authors", RFC 2223, October 1997.

GGF Copyright and Intellectual Property Guidelines

GGF has developed guidelines for copyrights for its documents and intellectual property guidelines. GGF requires a copyright notice for several reasons. First, the copyright gives GGF the right to publish the whole document as-is in perpetuity. Second, the copyright allows others to republish the whole document as-is without obtaining permission (e.g. a document repository or mirror site). Third, the copyright permits translation of the whole document into other languages. Finally, the copyright permits the development of derivative works within the GGF process. The authors retain all other rights. GGF intellectual property right management process is under development and is intended to mirror the intellectual property rights and procedures associated with the Internet Standards Process. In all matters of intellectual property rights and procedures, the intention is to benefit the Grid community and the public at large, while respecting the legitimate rights of others.

GGF has developed guidelines to contributing authors and the details can be obtained from their web site (http://www.gridforum.org/L_About/doc.htm)

GGF Membership

There are different levels of membership such as individual, general and for organizations and for sponsoring. The membership application and other details can be obtained by (http://www.gridforum.org/L_Involved_Mktg/member.htm) and the membership can be renewed annually.

GGF Working and Research Groups

The work of Global Grid Forum is performed within its various working groups and research groups. A **working group** is generally focused on a very specific technology or issue with the intention to develop one or more specific documents aimed generally at providing specifications, guidelines or recommendations. A

research group is often longer-term focused, intending to explore an area where it may be premature to develop specifications.

A *draft* document on GGF Structure **specifies** a set of "areas," each with a set of related working groups and/or research groups. These areas exist primarily for organizational and management purposes and is not intended to specify a comprehensive taxonomy for grids. It is also expected that new areas will be formed from time to time.
The groups listed below are currently active. Participation in groups can be electronic via mailing lists and/or by participating in person at GGF meetings, held three times per year.

GGF working groups and research groups produced over 100 draft documents prior to July 2001, some of which have been published in various journals. These are available at the group websites (http://www.gridforum.org/L_WG/wg.htm). The following table is reproduced from the GGF web site for information purposes only.

Table 1 GGF working groups and research groups

Area	Working Groups	Research Groups
Grid Information Services / Performance (GIS-PERF)	Grid Object Specification (GOS) Grid Notification Framework (GNF) Metacomputing Directory Services (MDS) Grid Monitoring Architecture (GMA) Network Measurement (NM) Discovery and Monitoring Event Description (DAMED)	Relational Database Information Services (RDIS)
Security (SEC)	Grid Security Infrastructure (GSI) Grid Certificate Policy (GCP)	
Scheduling (SCHED)	Distributed Resource Management Application API Working Group (DRMAA) Scheduling Dictionary (DICT) Scheduler Attributes (SA)	
Architecture (ARCH)	JINI	Grid Protocol

	NPI OGSI	Architecture (GPA) Accounting Models (ACCT)
Data	GridFTP Grid High-Performance Data Access and Integration Services	Data Replication (REPL) Persistent Archives (PA)
Applications, Programming Models, Environments (APME)		Applications and Testbeds (APPS) Grid User Services (GUS) Grid Computing Environments (GCE) Advanced Programming Models (APM) Advanced Collaborative Environments (ACE)
Peer-to-Peer*	NAT/Firewall Taxonomy Peer-to-Peer Security File Services Trusted Library	

*The Peer-to-Peer area was created in May 2002 and is in the process of being merged into GGF. All extant P2PWG working committees are working with the GFSG to transition from the now closed P2PWG into GGF.

Participation in GGF events

Participate in GGF Working Group or Research Group Activities
There are many diverse interests represented across the current GGF working groups and research groups.
One may participate in as many of Working Group (WG) or Research Group (RG) activities as interest to him or her. Participation can include any or all of the following:

1. Subscribe to the Group e-discussion list to keep abreast of, and to participate in, key discussion topics and document development.

2. Visit the WG, and relevant Area, web page to learn of developments, progress and meeting plans.

3. Contact the WG or RG Chair, or relevant Area Director(s), to express interest, offer feedback or seek out ways to get involved. If you are interested in offering help on a particular document, contact the document authors.

4. Participate in WG or RG sessions at the next GGF meeting.

Propose a *NEW* GGF Working Group or Research Group

GGF working groups and research groups are formed through an open process that concludes with formal approval of the group by the GGF Chair and the GGF Steering Group. The process for proposing a new WG or RG is as follows. In many cases it is advantageous to begin with a discussion with the GGF chair or a GGF steering group (GFSG) member to let them know what you are thinking of proposing. A proposed group may request an email distribution list to facilitate discussions and charter development.

1. Conduct a **Birds-of-a-Feather (BOF) session** for the next GGF workshop via email to the appropriate GFSG contact and to the Event Program Chair. (see **NOTE** below)

2. **Develop a draft charter** for the group, and then submit this draft charter to the GGF Chair. A *draft* document, "Global Grid Forum Structure," contains guidelines for developing a working group or research group charter as well as processes for working group or research group formation and operation.

3. Once the above steps have been followed and documented, send a formal request documenting the steps taken to the GGF Chair and the relevant Area Director(s). They will take the request to the GGF Steering Group. Nofitication is sent when formal recognition has been confirmed by WG or RG. ONLY formally approved GGF WGs and RGs may schedule WG/RG sessions during GGF Events.

NOTE- steps 1 and 2 may be done in any order.

GRID tools and technologies

This is not a comprehensive list of all the available technologies related to Grid computing. Such a list could be found from GGF and gridcomputing web sites listed earlier. The following is brief list of some important software tools that are widely used in the Grid community

The SDSC *Storage Resource Broker (SRB)* is client-server middleware that provides a uniform interface for connecting to heterogeneous data resources over a network and accessing replicated data sets. SRB, in conjunction with the Metadata Catalog (MCAT), provides a way to access data sets and resources based on their attributes rather than their names or physical locations. Data resides in three forms: database, Unitree, HPSS. (http://www.npaci.edu/DICE/SRB/)

The *Globus Project* is developing fundamental technologies needed to build computational grids. Grids are persistent environments that enable software applications to integrate instruments, displays, computational and information resources that are managed by diverse organizations in widespread locations. (http://www.globus.org)

GRID FTP – Incorporates FTP, involves sub-setting, secure data transfer,

Issues:

Security, data authentication, applications

Data cutter technology (http://www.cs.umd.edu/projects/hpsl/Research-Areas/DataCutter.htm)

AMES IPG power GRID activities (http://www.npaci.edu/envision/v16.2/power-grid.html)

GRID in a Box CD-ROM (GRID in a Box, Cluster in a Box, Access in a Box). The Alliance's Grid-in-a-Box (GiB) is a set of software tools that lower the barriers to getting on the grid, using grid resources, and offering resources. The grid infrastructure, with the GiB software as an integral part, will link advanced supercomputers and computing clusters, visualization tools, and mass storage devices via high-speed networks into a powerful, flexible problem-solving environment. GiB includes middleware for tasks such as authentication, job management, and information discovery. By simplifying these tasks, GiB will help to expand the use of the grid more rapidly. (http://www.ncsa.uiuc.edu/Tech-Focus/Deployment/GiB/)

Reference:

Grid Services for Distributed System Integration , IEEE proceedings, 2002, I. Foster, C.Kesselman, J.Nick, and S. Tuecke.

3.4 Sun's Java Community Process

The Java Community Process (JCP) was developed to produce and revise high quality Java specifications in "internet time" using a consensus building approach. The JCP is composed of four major steps, depicted in Figure 1 :

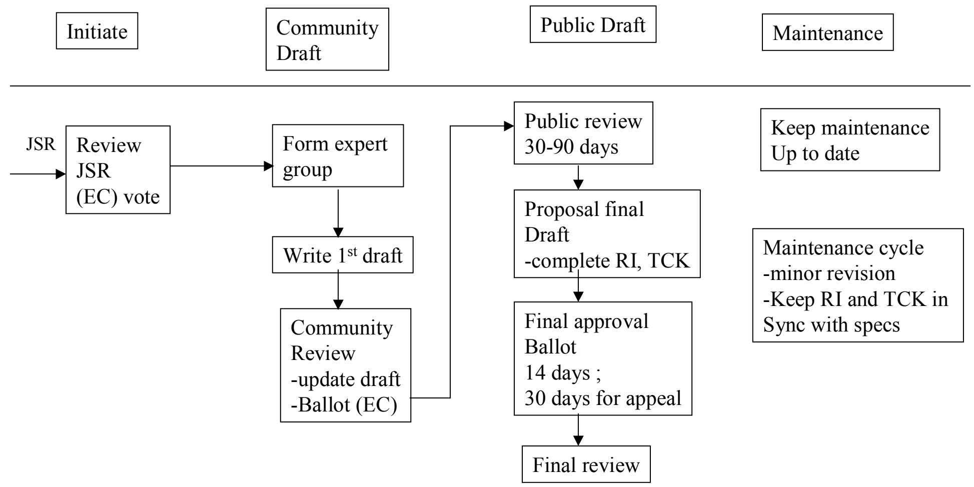

Figure 1. Java Community Process

1) Initiate: One or more community members can submit a Java Specification Request (JSR) to the Program Management Office (PMO) for development or significant revision of existing specifications. The Executive Committee reviews the JSR, considers comments from the community, and votes on the approval of the JSR.
2) Community Draft: A group of experts develops the first draft and refines it based on comments from the community. The Executive committee votes to determine whether the draft should proceed to public review.
3) Public draft: the public reviews the Internet-accessible draft during a 30-90 day period. The group lead completes a reference implementation (RI) (a prototype / proof of concept) and Technology Compatibility Kit (TCK) (a suite of tests). Based on comments received, final draft is prepared from comments received and voted by the Executive committee for final approval.
4) Maintenance: To tackle the on-going requests for clarifications, enhancements and revisions even after the final draft is approved, a maintenance lead is assigned.

Definitions

Program Management Office (PMO): This group (SUN) is responsible for administering the JCP and chairing the Executive Committee

Executive Committee (EC): Comprised of major stakeholders and members of the Java community, this group is responsible for approving specifications and reconciling discrepancies

Reference Implementation (RI): Prototype / proof-of-concept of a specification

Technology Compatibility Kit (TCK): A suite of tests to determine whether an implementation complies with the specification.

Maintenance Lead (ML): Expert responsible for maintaining the specification

4 EXPERIENCES OF STANDARDS ORGANIZATIONS

As a counterpart to NASA-ESE's own experience with standards design, adoption, and deployment, the next several sections examine the processes and outcomes of several standards organizations working in related topic areas:

- The International Organization for Standardization (ISO) and its Technical Committee 211 on Geographic information/Geomatics;
- The Open GIS Consortium (OGC);
- The World Wide Web Consortium (W3C);
- The Consultative Committee for Space Data Systems (CCSDS);
- The U.S. Federal Geographic Data Committee (FGDC); and
- The Internet Engineering Task Force (IETF).

Topics of interest in each case include not only the standards of relevance to ESE / SEEDS, but also the decision processes followed by these groups, and the varieties of success that they have experienced.

4.1 ISO TC 211 Geographic information/Geomatics[3]

4.1.1 Description

ISO's Technical Committee (TC) 211 on Geographic Information / Geomatics aims to establish a structured set of standards for geospatial information, including methods, tools, and services for defining and describing data, and acquiring, processing, analyzing, accessing, presenting and transferring such data in digital / electronic form among different users, systems and locations.

TC 211, established in 1994, brings together 53 national bodies (members) and 17 liaison organizations. TC211 is organized as Working Groups (including project teams and editing committees) and Advisory Groups. The TC assigns each new Work Item to a Working Group, which establishes Editing Committees to disposition comments on a specific draft standard. Advisory Groups develop recommendations to the TC.

4.1.2 Standards Relevant to ES

NCITS panel L1[4], as chartered by ANSI, represents the US in TC 211: it formulates US positions and organizes US member votes on ISO TC211 issues. NCITS L1 also decides which TC211 standards are to be adopted as ANSI

[3] http://www.statkart.no/isotc211/

[4] http://gsvaresa07.er.usgs.gov/QuickPlace/ncits_l1/Main.nsf

standards. Table 2 below lists TC211 standards that NASA considers a high priority for adoption as US standards.

Table 2 . ISO Standards relevant to NASA

ISO Number	Title	ISO Status	Abstract vs. implementation
19104	Terminology	FDIS pending	Abstract
19107	Spatial schema	FDIS	
19108	Temporal schema	FDIS	Abstract
19111	Spatial referencing by coordinates	FDIS	Abstract
19112	Spatial referencing by geographic identifiers	DIS	Abstract
19115	Metadata	DIS	Both
19119	Services	DIS pending	Abstract
19123	Schema for coverage geometry and functions	CD2	Abstract
19127	Geodetic codes and parameters	Open WI	Abstract
19128	Web map server interface	CD1	Implementation
19129	Imagery, gridded and coverage data framework	Open WI	Abstract
19130	Sensor and data model for imagery and gridded data	Open WI	Abstract

DIS: Draft International Standard
FDIS: Final Draft International Standard
CD: Committee Draft
WI: Work Item

4.1.3 Standard Work in Progress

The third column of Table 2 indicates the status of the standards in work in TC211.

4.1.4 Standard Processes

ISO/IEC Directives[5] define processes for standards development in TC211, including the stages that a document must go through to become an ISO Standard:

- Preliminary work item (PWI)
- New work item proposal (NWIP or NP)
- Working draft(s) (WD) – may be omitted
- Committee draft(s) (CD) – may be omitted
- Enquiry draft – ISO Draft International Standard (DIS), IEC Committee Draft for Vote (CDV)

[5] http://www.iso.ch/sdis/directives

- Final draft International Standard (FDIS) – may be omitted
- International Standard -ISO, IEC or both ISO/IEC

A project team in a working group develops a Working Draft. The project team consists of a project leader, an editor, and a group of experts nominated by TC 211 national bodies. Once the working group releases a Committee Draft, the entire TC may comment on the document; and an Editing Committee formed specifically for that CD deals with the comments. National bodies and liaison organizations submit comments on the CD using a TC211 template. The Editing Committee answers each of the comments. TC211 has, on average, developed two CD's before considering advancing the document to DIS. Once the editor drafts a DIS, the TC reviews it for objections to passing it to the ISO Central Secretariat for circulation as a DIS.

4.1.5 Current Status

ISO TC 211 began in 1994 to develop a comprehensive and interrelated set of standards covering the geographic information domain. This was an ambitious task building on many existing national and multi-national standards. TC211 is presently issuing the first generation of these standards as Final Draft International Standards (FDIS). There are many ways to assess this endeavor:

- Was the goal of comprehensive standards met? The sheer number of standards that TC211 is producing is overwhelming to the casual observer, but this is not a measure. There can be no doubt that the domain of geographic information cannot be fully comprehended by a single individual. After the initial slate of standards (19101 to 19120) was started, it was realized that additional topics were needed.
- Was the goal of interrelated standards met? The initial approach was that all of the standards would be parts of a single ISO standard; e.g., originally ISO 19102 was ISO 15046-2. Some participants viewed the change to the 19100 numbering scheme as the demise of the interrelationship. But the 19100 series continues to remain interrelated. This element is good because it ensures consistency (e.g., all standards point to 19111 for coordinate reference systems). The interrelationship is bad, as you may need several of the items to implement the standard.
- Is the emphasis on abstract standards appropriate? The approach of ISO TC211 is to define the concepts of geographic information and a process for implementing the concepts. Only one of the first twenty TC211 standards addresses bits-on-the-wire (ISO 19118); and only two (19115 and 19119) suggest actual XML encoding structures. The assumption here is that the concepts of geographic information remain fairly constant in comparison to the rate of change of software implementation mechanisms.
- What is the quality of the standards? Measured against the goals of TC211, the quality is undoubtedly world-class. Most of the work in TC211 is based on

previously implemented standards, which have been merged into single international standards by some of the world's experts in the field.

- Can I directly implement the TC211 standards? In most cases the answer is no. The TC211 standards provide the concepts that system designers, data engineers, and software developers can use to develop implementation-level standards. In response to these concepts, TC211 and the Open GIS Consortium (OGC) developed a cooperative agreement. OGC focuses on implementation specifications that implementers can use directly to achieve interoperability. TC211 is adopting implementation specifications developed by OGC; and conversely, OGC is adopting the abstract specifications developed in conjunction with TC211.

Success Criteria	**ISO TC211 Standard**	**Comments**
Is the development of standards process open and transparent?	Yes	Any organization can participate through their national standards body or as an affiliate member
Are the standards developed by this organization freely available - open standards?	Yes	All ISO standards are published and available to anyone. There is a publication charge for most of the standards
How much time did it take for the development of the standard- months, years?	2-5 years	Depends upon the maturity at the start of the project. ISO rules require re-balloting the project if it takes longer than 5 years.
Does the standard have software tools support?	Not a requirement	Depending upon the standard, software may not be applicable, e.g., abstract, architecture, data model, terminology specifications. Implementation standards adopted by TC211 have software support.
Vendor support to the standard	Not a requirement	Depending upon the standard, vendor support may not be applicable, e.g., abstract, architecture, data model, terminology specifications. Implementation standards adopted by TC211 have vendor support.
Relationship with other standards organizations	Excellent	ISO has liaison agreements with 18 external and 11 internal (ISO/IEC TC/SC) standards organizations. A cooperative agreement exists between TC211 and OGC.
Has the organization achieved its stated goals?	Yes	ISO TC211 is finishing its first standards, which began as a coordinated set. It is too early to determine the success of individual standards. As an organization, TC211 has become recognized in the geographic information field, i.e., beyond GIS software.
Usage of standards 1. How widely are they used? 2. ESE domain	Too early to judge	Some of the standards that TC211 standards are based on are well established, e.g., DIGEST and FGDC standards.

Success Criteria	ISO TC211 Standard	Comments
Quality 1. Documentation and description of the standard 2. Multi-language support 3. Quality of the standard test suite	1. Excellent 2. English is required, some also in French. 3. Excellent	1. In addition to the body of the standard, ISO requires terminology and conformance sections. 2. TC211 has a specific focus on cultural and linguistic adapability
Performance: 1. Usage 2. Adoption 3. Engineering and integration	Too early to tell	
Ease of use	Varies	
Evolution Is there a process for this?	Yes	By corrections or amendments
Metadata support Interoperability	Excellent	ISO 19115 is a draft international standard developed from FGDC metadata standard. Planning has begun to adopt 19115 as an ANSI standard and as a revision of FGDC metadata.
Are these standards relevant to ESE?	Highly relevant	Required by federal directives
NASA participation	NASA ESE	

4.1.6 NASA Current Involvement

The ESE provides most of the NASA involvement in TC211. NASA's representatives to NCITS L1 are from ESE, as are most of the experts who have participated in the development of TC211 standards. ESE provides project leaders, editors and experts for several of the completed and current TC211 standards.

U.S. public laws require NASA to use relevant consensus-based international standards. World Trade Organization agreements also require government procurements to follow international standards. ISO is a recognized international standard-setting organization. Therefore, NASA's involvement in development of relevant ISO standards is not only important to SEEDS but also to current NASA ESE missions.

4.2 Open GIS Consortium (OGC)

4.2.1 Description

http://www.opengis.org

The Open GIS Consortium is an international consensus-based organization with the following vision:

> *"OGC envisions the full integration of geospatial data and geoprocessing resources into mainstream computing and the widespread use of interoperable, commercial geoprocessing software throughout the global information infrastructure."*[6]

OGC is organized as a not-for-profit membership organization and was founded in 1994. Members include universities, small, medium and large commercial enterprises, and government organizations. OGC is headquartered in the United States (Wayland, Massachusetts) but functions as a virtual organization with staff members in MA, IN, VA, and Europe. Meetings are held five times per year (typically the first full week of the even months except August) and are hosted by members around the world. OGC has a multilevel structure reflecting various levels of membership fees. Membership levels are University ($300), Associate ($4000), Technical ($10,000), Principal ($50,000), and Strategic (negotiated). Only Technical and higher-ranked members may vote on the adoption of specifications or the release of documents outside of OGC. At the working group level, any member may vote.

OGC maintains close relationships with ISO TC211 (section 4.1), and is pursuing a strategy of "co-branding" standards with ISO TC211. OGC also maintains relationships with other standards bodies at varying levels, ranging from cooperation to cross-membership.

OGC has developed a commanding position in the area of geospatial standards. Government organizations are beginning to require OGC standards compliance in procurements, OGC is mentioned in many conferences dealing with spatial topics, and OGC is taken quite seriously by the international spatial community. At the same time, OGC and spatial standards are not well known outside of that community. When organizations such as the Internet Engineering Task Force (IETF) (section 4.6) touch upon spatial topics, their members do not generally know about OGC. This may slowly be changing as awareness of OGC grows.

OGC also operates an Interoperability Program (IP), which consists of a series of funded initiatives that are designed to provide fast-track development of interfaces and encodings that have a high potential for being adopted as OGC specifications. The IP began in 1999 as the Web Mapping Testbed and has expanded greatly since then, encompassing testbeds and pilot projects. Testbeds are meant to generate new interfaces and encodings or to improve upon old ones in a substantial way, while pilots are meant to test interfaces and encodings in a more operationally focused way (a NASA prototype would be an OGC pilot).

---

[6] Open GIS Consortium, Inc., *The Open GIS® Guide*, Third Edition, June 3, 1988

4.2.2 Standards Relevant to ES

OGC has developed a number of standards, most of which are directly relevant to ES. Furthermore, the OGC "Abstract Specification" provides a broad-reaching statement on what areas will ultimately be tackled by OGC and where standards may be forthcoming. OGC's interface and encoding documents fall into three major categories:

- Specifications are definitive statements about an interface or encoding. They "are the result of OGC's Technology Development Process and are engineering specifications that implement part of the Abstract Specification for particular distributed computing platforms.[7]"
- Recommendations are published when there is considerable agreement among the members and "are the official position of OGC, but are not Adopted Specifications. If Recommendations are engineering specifications, then one may assume that the content is still under rapid change and that the reader should recognize the volatile nature of the specification when building implementations.[8]"
- Discussion papers "present a discussion of technology issues considered in the Working and Special Interest Groups of the Open GIS Consortium Technical Committee. The content of these papers is presented to create discussion in the geospatial information industry on a specific topic; the content of these papers is not to be considered an adopted standard of any kind. These papers do not represent the official position of the Open GIS Consortium nor of the OGC Technical Committee.[9]"

[7] source: OGC web site [http://www.opengis.org/techno/specs.htm]

[8] source: OGC web site [http://www.opengis.org/techno/specs.htm]

[9] source: OGC web site [http://www.opengis.org/techno/oip.htm#discussions]

Table 3 OGC Specifications, Recommendations, and Discussion Papers of high relevance to NASA-ESE

OpenGIS® Catalog Interface Implementation Specification	1.0	1999-08-02	**Very important** to ESE, especially the Stateless Catalog work derived from this.
OpenGIS® Web Map Server Interfaces Implementation Specification	1.0.0	2000-04-19	V**ery important** to ESE, even though later versions exist.
OpenGIS® Web Map Server Interfaces Implementation Specification	1.1.1	2001-12-10	V**ery important** to ESE. It adds some features that were lacking in the 1.0 version.
OpenGIS® Geography Markup Language (GML) Implementation Specification	2.1.1	2002-04-11	**Important** to ESE. It provides a basis for expressing vector data.
OpenGIS® Recommendation - Recommended Definition Data for Coordinate Reference Systems and Coordinate Transformations	1.0.1	2001-04-05	**Important** to ESE.
OpenGIS Discussion Paper 01-044r2: Units of Measure and Quantity Datatypes		2001-06-15	**Important** for ES. It is being folded into GML v3.0.
OpenGIS Discussion Paper 01-036: Gazetteer Service Specification	0.084	2001-03-15	R**elevant** to ES. Vendors should be encouraged to develop it further.
OpenGIS Discussion Paper 01-035: Geoparser Service Specification	0.7.1	2001-03-27	**Interesting** and has potential for ES.
OpenGIS Discussion Paper 01-026r1: "Geocoder Service Draft Candidate Implementation Specification 0.7.6"	0.7.6	2001-03-28	In**teresting** to ES. Vendors should be encouraged to develop it further.
OpenGIS Discussion Paper 01-018: Web Coverage Server	0.4	2001-02-06	V**ery important** to ES. NASA should encourage its developers to follow through and turn it into an OGC specification. It is also under further development in the OWS testbed.
OpenGIS Discussion Paper 01-023: Web Feature Server	0.0.12	2001-01-21	**Important** to ES. NASA should encourage its developers to follow through and turn it into an OGC specification. It is also under further development in the OWS testbed.
OpenGIS Discussion Paper 01-028: Styled Layer Descriptor	0.7.0	2001-02-07	Material in this paper is **important** to ES.
OpenGIS Discussion Paper 01-019: XML for Imagery and Map Annotations	0.4	2001-02-06	Material in this paper is **important** to ES.

OpenGIS Discussion Paper 01-013r1: High-Level Ground Coordinate Transformation Interface	none	2001-02-27	This paper contains an **important** design idea for converting fine-grained interfaces into higher-level coarse-grained interfaces that can be used in Web services.
OpenGIS Discussion Paper 01-024r1: Web Registry Server	0.0.2	2001-01-26	**Important** paper for ES. Combined with the Catalog Services specification, it forms the basis for an eventual stateless catalog specification.
-- placeholder for terrain view server --			

4.2.3 Standard Work in Progress

There are two major areas of work in progress. Within the OGC Specification Program, there is currently a Request For Proposal (RFP) for “OpenGIS Feature Geometry”. This RFP is based on the content of ISO 19107 “Spatial Schema.” Initial responses to this RFP are due in December 2001 at the Vancouver, Canada meetings of the OGC. Within the OGC Interoperability Program, the OGC Web Services (OWS) testbed and Open Location Services (OLS) testbed are in progress. The most recent phase of the OWS Testbed worked on the following topics:

- Common Architecture – This work included a Service Model for spatial Web services, catalog services, and registry services. (Very important to NASA ESE.)
- Web Mapping – This work included further developments of the Web Map Server (WMS), Web Feature Server (WFS), and Web Coverage Server (WCS) specifications. (Very important to NASA ESE.)
- Sensor Web Enablement – This work included encodings of sensor models (primarily *in-situ* sensors such as water quality sensors) as well as some initial interfaces for sensor services. This thread is important to NASA ESE insofar as it can be integrated into more traditional remote sensing models and architectures. It is also important since cal-val activities rely on *in-situ* sensors.

4.2.4 Standard Processes

4.2.4.1 OGC Structure

OGC’s Board of Directors provides oversight and works through OGC staff to influence the consortium’s strategic direction. The Planning Committee (PC) (comprised of Principal and Strategic Members) makes final decisions on releasing specifications and provides some guidance on Interoperability Program initiatives. The Technical Committee (TC) makes decisions on how documents are packaged and released (pending PC approval). Within the Technical Committee, Special Interest Groups (SIG) and Working Groups (WG) define engineering priorities and develop specifications. Members at any level may vote in the SIGs and WGs.

OGC has developed two distinct modes of operation. One is the Specification Program, where Working Groups typically design requirements for interface specifications; OGC releases these as a call for implementation-based responses. Those responses are then evaluated and possibly turned into OGC specifications. More details on this process are provided below.

The other mode, a more recent development, is the Interoperability Program, in which sponsors fund a group of vendors and R&D groups to build prototype implementations of interfaces in a rapid-cycle, iterative process, and to derive interface descriptions from the lessons learned in the development. These interface descriptions are then evaluated and possibly turned into OGC specifications. More details on this process are provided below.

OGC's structure and processes can be complex. In fact, they are modeled on the OMG (Object Management Group), one of the first industry consortia to produce implementation specifications successfully. The process is designed to satisfy US anti-trust laws governing the development of technology by groups of commercial enterprises.

4.2.4.2 Specification Program (SP)

Figure 2 shows the process flow found in the SP.

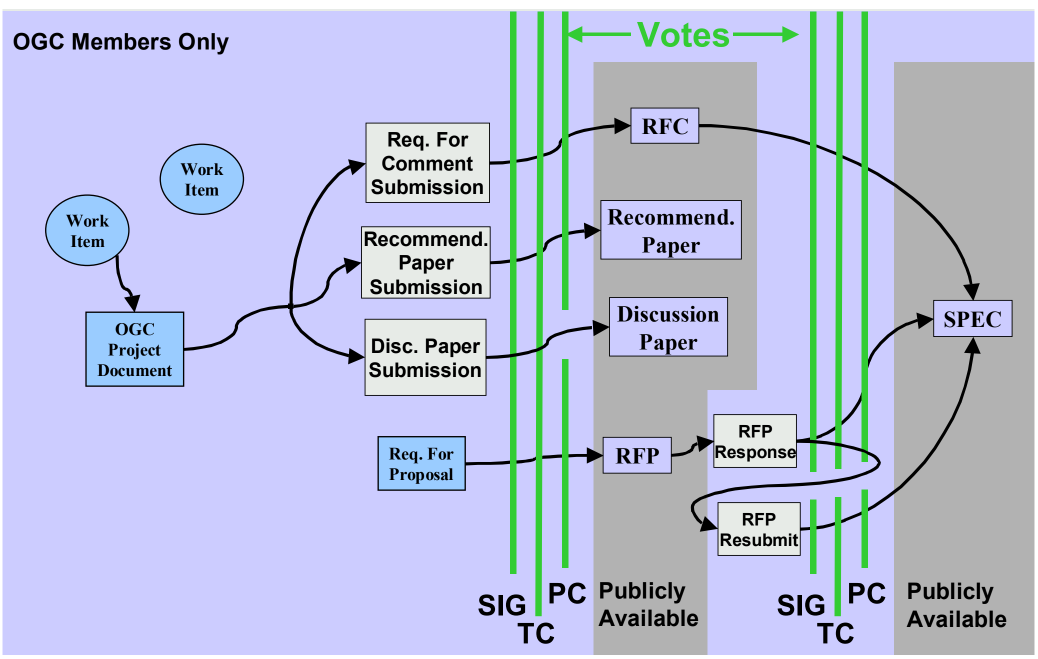

Figure 2 OGC Specification Program Process

OGC members work in Special Interest Groups (SIG) or Working Groups (WG) to describe particular interfaces or encodings; they then decide whether the resulting document should be turned into a Request For Proposal (RFP). Alternatively, OGC SIGs, WGs, or individual members can submit a Request For Comment (RFC), a Recommendation Paper, or a Discussion Paper. In all cases, the SIG or WG and the Technical Committee must approve the public release of any document; the Planning Committee must also approve the public release of all but Discussion Papers.

RFPs then follow a track that involves industry evaluation of the RFP, submission of responses to the RFP, and evaluation of the responses by the OGC SIG or WG. The SIG or WG can then either approve a response or request a resubmission. RFP submissions tend to be made by teams of companies and the resubmission process is designed to allow competing teams to consider joining forces rather than face a second rejection. Approved RFP submissions then become OGC specifications.

Approved RFC submissions are released to the public for a comment period, after which the RFC submitter(s) may adjust the content based on the comments. The revised content runs another gauntlet of votes before becoming an OGC specification. In one case (GML 1.0), the Technical Committee turned an RFC submission into a Recommendation Paper. GML 2.0 was then produced via a revision process (which is not shown here).

4.2.4.3 Interoperability Program

Figure 3 below shows the IP process.

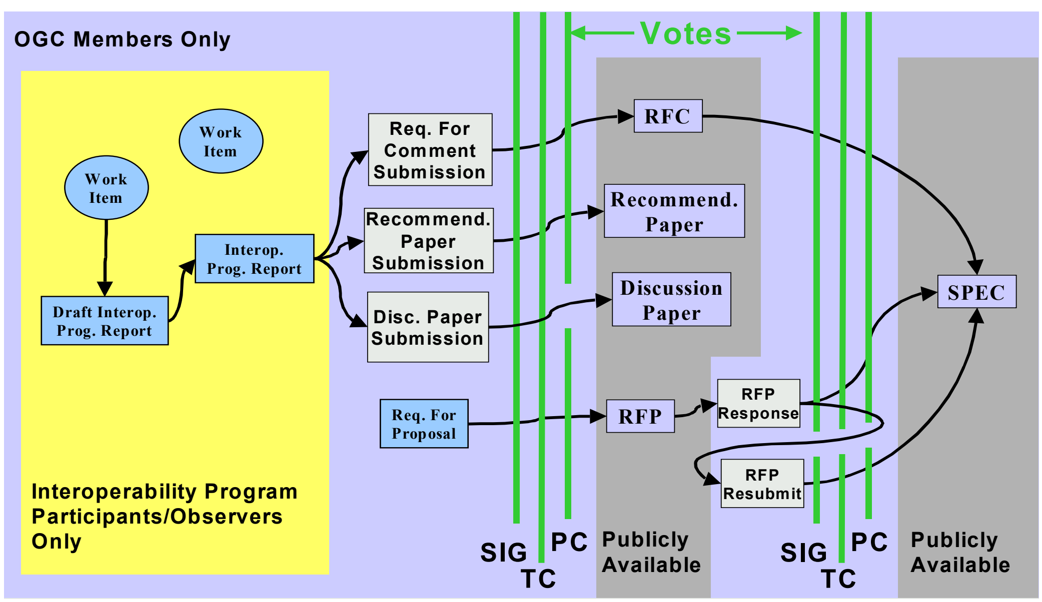

Figure 3 OGC Interoperability Program Process

The net result of the Interoperability Program (IP) is quite similar to that of the Specification Program (SP). One major difference is that while the SP formalizes mostly matured technologies into OGC specifications, the IP experiments with new technologies of interest to the GIS community, in a shared testbed. In this testbed, prototype implementations build a consensus and support an IP report. The IP report must still go through the SP process to become an OGC Spec; however, the IP process can respond more rapidly to changing technology, and lets vendors build interoperable products in advance of an OGC specification.

IP participants develop “Draft Interoperability Program Reports” (or “Interoperability Program Reports”). These are then available for use in the Specification Program as material for RFC submissions, Discussion Papers, etc. Most of the current OGC Discussion Papers are the result of Interoperability Program work.

4.2.5 Current Status

To date OGC’s results are mixed. OGC has been spectacularly successful in gathering the vendor community, the academic community, and the end-user community (including NASA) into a fairly coherently aligned body. The entire spatial community seems to share a frame of reference in the area of specifications and the underlying abstract concepts as a result of participation in OGC or influence by OGC. In the last year, published papers have begun to mention OGC specs and ideas either because the authors have implemented or tested them, or because the authors feel it’s important to align with the OGC. Government agencies in the US and abroad are beginning to require OGC specifications (or at least OGC-awareness) in their procurements.

In terms of the adoption of specifications and their commercial success, the Specification Program has been a slow producer. The SP took nearly three years to produce its first specification (“Simple Features”) and although a few vendors implement this specification in their products today, none can claim a commercial benefit for the vendor, or an operational benefit to end users. Other specifications developed entirely within the SP are similarly under-subscribed, with one notable exception – the OGC Catalog Services Specification, co-sponsored by NASA, is in use in several locations worldwide.

The IP has produced three specifications to date: the Web Map Service (WMS) , Geography Markup Language (GML), and Web Feature Service (WFS). WMS took about six months to develop within the IP, then two additional months to clean up and turn into a usable document, and then about four months to get approved by the SP. Before it was approved, OGC member companies had begun developing products based on it; and shortly after its adoption they were deploying commercial implementations. GML took a little longer; its first version was issued as a Recommendation by the SP and only became a specification in its second revision. GML thus took two years to become a spec. GML has generated intense interest and the British Ordnance Survey (British equivalent of the USGS) is now distributing its Digital National Framework[10] data in GML.

The IP has produced several other interface descriptions that have been published as discussion papers, but not yet endorsed by the SP. This is due primarily to two factors. First, some SP participants view the IP with some suspicion as "hackers", and second, many IP participants are not as "savvy" as they need to be. For instance, the Web Feature Server languished for over six months after being approved for release as an RFC because the WFS submitters were not able to continue working the process. Yet companies were building WFS implementations based on the discussion paper.

4.2.6 NASA Current Involvement

NASA has been an OGC supporter since its inception in 1994. This support intensified when NASA co-sponsored the OGC Catalog Services Specification and the Interoperability Program (known then as the Web Mapping Testbed). NASA currently sponsors the Open Web Services initiative of the Interoperability Program and supports the Earth Observation SIG. NASA is a Principal Member of OGC and has a vote at the Technical Committee level as well as at the Planning Committee level.

4.3 The World Wide Web Consortium (W3C)

4.3.1 Description

The World Wide Web Consortium (W3C) is an international organization created in October 1994 to advance the evolution and interoperability of the Web through common protocols. W3C has over 510 member organizations and 60 full-time staff from around the world and it has earned international recognition for its contributions to the growth of the Web. W3C develops interoperable technologies (specifications, guidelines, software, and tools) to "lead the Web to its full potential" as a forum for information, commerce, communication, and collective understanding. Many of the W3C standards such as HTML, XML, MathML, and SVG have been widely used in scientific applications including Earth Science.

As a vendor-neutral organization, W3C promotes interoperability by designing and promoting open (non-proprietary) computer languages and protocols that avoid market fragmentation. This is achieved through developing industry consensus and encouraging an open forum for discussion.

Membership in W3C is open to all types of organizations: commercial, educational, and governmental entities; for-profit or not-for-profit organizations; with two classes of membership: Full and Affiliate. There are no differences in benefits between the two classes of membership.

[10] http://www.ordnancesurvey.co.uk/downloads/mm/osmm_in_gml_format.PDF

W3C organizes the work necessary for the development or evolution of a Web technology into Activities. Each Activity has its own structure, but an Activity typically consists of one or more Working Groups, Interest Groups, and Coordination Groups. Within the framework of an Activity, these groups generally produce Recommendations and other technical reports as well as sample code.

Working groups are small groups (less than 15 people) of experts responsible for deliverables. Interest groups are groups of people interested in evaluating and discussing Web technologies and ideas. The coordination group's role is to facilitate communication between groups (internal and external). Each activity is also grouped into five domains: Architecture, Document Formats, Interaction, Technology and Society, and Web Accessibility. Within an activity, groups produce recommendations and other technical reports. The overall structure is illustrated below in Figure 4.

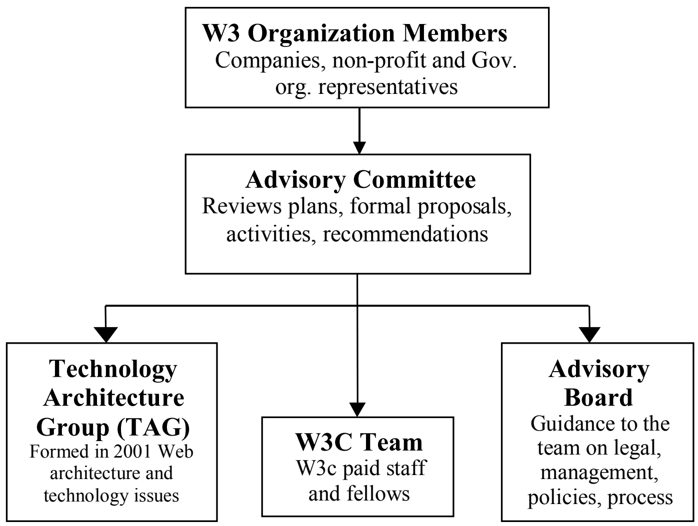

Figure 4 W3C Organizational structure

4.3.2 Standard Processes

W3C publishes two types of reports: Recommendation track Technical Reports and W3C Notes. To become a standard, a Technical Report goes through the following stages.

Working draft (WD): a technical work in progress

Last call working draft: a draft released to the public for review and comments

Candidate recommendation (CR): a call for implementation experience outside the working groups and W3C.

Proposed recommendation (PR): a report on dependencies

Recommendation (REC): a proposed standard

Figure 5 shows the steps in the development of standards along the Recommendation track.

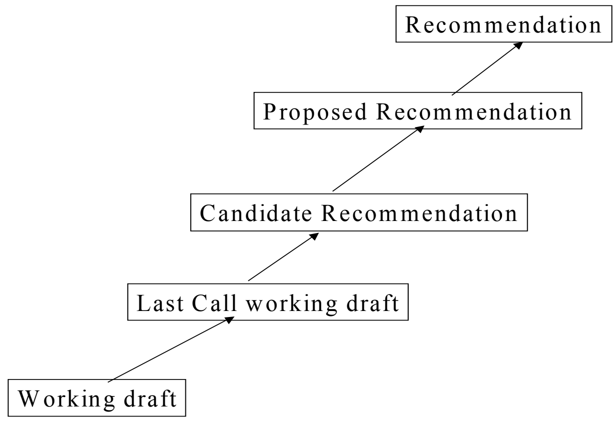

Figure 5 The W3C Recommendation track

The recommendation is the final step in this process. A W3C Note does not follow this track; rather, it is a dated public record of an idea or comment, published at the discretion of the W3C director.

The process is based on consensus and open exchange of information. However, the W3C director has the final authority to promote a document to a recommendation.

4.3.3 Standards relevant to ES

Table 4 describes the W3C standards relevant to ESE and their status.

Table 4 W3C Standards Relevant to ESE

Name	Description	status
XML 1.0	Extensible Mark up Language http://www.w3.org/XML/	Recommendation
XSLT	XSLT transformation http://www.w3.org/Style/XSL/	Recommendation
XML schema	XML schema http://www.w3.org/XML/Schema	Recommendation
Xlink	XML linking language http://www.w3.org/XML/Linking	Recommendation
XHTML		Working Draft
Xforms	http://www.w3.org/MarkUp/Forms/	Working Draft
Xpath	http://www.w3.org/TR/xpath	Recommendation
Xpointer	XML Pointer Language http://www.w3.org/XML/Linking	Last Working Draft

Xbase		Recommendation
MatML 2.0	Mathematical Markup Language http://www.w3.org/Math/	Recommendation 2/21/2001
HTML		Recommendation
HTTP and HTTP/1.1	Hypertext Transfer Protocol. RFC 2617: HTTP Authentication (Feb 1999) RFC 2774: HTTP extension framework (Feb 2000) http://www.w3.org/Protocols/ http://www.w3.org/Protocols/Specs.html	HTTP activity closed Allocated to jigsaw activity
SMIL	Synchronized Multimedia Integration Language http://www.w3.org/AudioVideo/	Recommendation
RDF	The Resource Description Framework http://www.w3.org/RDF/	Recommendation
SOAP	Simple Object Access Protocol http://www.w3.org/2000/xp/	Working Draft
PNG	Portable Network Graphics http://www.w3.org/Graphics/PNG/	Recommendation 10/1/1996
SVG 1.0	Scalable Vector Graphics http://www.w3.org/Graphics/SVG/Overview.htm8#news	Recommendation 9/5/2001
URI/URL	Uniform Resource Identifiers http://www.w3.org/Addressing/ Uniform Resource Locator	

4.3.4 Current Status

W3C began in 1994 to develop a comprehensive and interrelated set of standards covering all aspects of the Web, to promote its evolution and ensure interoperability. Many of the standards and protocols developed by W3C are already in use. There are many ways to assess this endeavor.

The sheer number of standards that W3C is producing is overwhelming to the casual observer, but this is not a measure of success or failure of the process or organization. Some of the W3C standards, like HTML and HTTP, have gained universal acceptance and many others relating to the semantic Web are still under development. W3C has made highly significant progress towards meeting the goals of development of Web infrastructure and ensuring interoperability. Table 5 gives a description of some measures by means of which the success of a standard can be measured.

The success of a standard should be measured against its usage and acceptance in the community and industry. W3C gets outstanding ranking in this area. Quality and performance are other issues that can be used to measure the success of the standard. W3C gets a very high ranking in this area. The other criteria of measure should be interoperability. W3C standards help the seamless interchange and distribution of information through web and gets high ranking.

Some of the issues to be considered in recognizing the success or failure of standards promoted by the W3C organization include its relatively short history of its organization as a standards body and the fact that many of the W3C standards are under development and are evolving. Some examples of highly successful W3C standards are HTML and HTTP. Although XML has achieved a great degree of success, the semantic Web has yet to be realized.

Table 5 Success Criteria of W3C Standards

Success Criteria	W3C Standard	Comments
Is the standards development process open and trans-parent?	Yes	Any organization can partic-ipate by becoming a member or affiliate member
Are the standards developed by this organization freely available - open standards?	Yes	All W3C standards are pub-lished and freely available
How much time did it take for the development of the stan-dard- months, years?	Generally most of the W3C standards become recommen-dations in a few months to a year	The appendix list of this document gives the details
Does the standard have soft-ware tools support?	Yes	All the W3C standards have one or more public domain tools to support and test the standard
Vendor support to the standard	Excellent	Many commercial companies are members of W3C and therefore there is very good support in terms of COTS
Relationship with other stan-dards organizations	Excellent	W3C works with many other standards organizations such as OGC, ISO and IEEE
Has the organization achieved its stated goals?	Yes in many areas the pro-gress is excellent, but there is still work in progress in some areas such as semantic Web	
Usage of standards 1. How widely are they used? 2. ESE domain	1. Some of the W3C stan-dards, such as HTML, are used universally. In general, most of the W3 standards are used widely by many different organizations 2. Some of the W3C standards are widely used in the ESE domain such as HTML, HTTP and XML	With so many W3C standards, it is difficult to come up with a simple statement. But in gen-eral the Web itself is used uni-versally and so also are many of the W3C standards. Some of the W3C standards are works in progress
Quality 1. Documentation and descrip-tion of the standard	1. Excellent	

2. Multi-language support 3. Quality of the standard test suite	2. Excellent 3. Excellent	
Performance: measured in terms of 1. Usage 2. Adoption 3. Engineering and integration	1. Excellent 2. Excellent 3. Excellent	
Ease of use	Most of the W3C standards are very easy (HTML, XML, HTTP) to use by both end users and system engineers, developers, and other technical people	
Evolution Is there a process for this?	Yes	
Metadata support Interoperability	Very good Excellent	
Are these standards relevant to ESE?	Highly relevant	
NASA participation	AMES Research Center is an official W3C member. Many in the NASA community are either directly or indirectly involved in the development process.	

4.3.5 Lessons Learned

The standards development process should be based on consensus, industry involvement and open exchange of information. W3C revolutionized the process of standards development. For the first time, it recognized that standards should evolve with technology and software tools. Standards are not mere specification documents, and have to be supported with software tools and adequate testing. Standards should also be developed in a relatively short period of time.

4.3.6 NASA Current Involvement

NASA, Ames Research Center is an official member of the W3C organization. Many NASA-funded projects and organizations participate and use standards promoted by W3C.

4.4 The Consultative Committee for Space Data Systems (CCSDS)

4.4.1 Description

The Consultative Committee for Space Data Systems (CCSDS) is an international organization of space agencies and industrial associates interested in developing standard data handling techniques to support space research.

CCSDS was formed in 1982 by space agencies and is composed of the following:

- 10 member agencies that fully participate in the CCSDS activities.
- 23 observer agencies that participate at a reduced level of effort
- 100 industrial associates (and liaison organizations). Associates are scientific or industrial organizations that monitor closely the technical document development process. Liaison organizations are organizations with space related programs.

Some of the goals of CCSDS are to:

- Reduce space mission costs
- Improve cross support and space data understanding
- Ensure space data preservation

Further information on CCSDS can be found at http://www.ccsds.org/

4.4.2 Standard Processes

The CCSDS standard process is composed of the following books:

- Blue books are CCSDS recommendations
- Red books are issues of the draft recommendations that are undergoing formal reviews. It may take more than one Red Book review cycle to get a Recommendation to Blue Book (final CCSDS standard) status.
- Pink books are proposed updates to current recommendations (i.e., to existing Blue Books)
- Green books are CCSDS reports
- Yellow books are CCSDS administrative reports
- White books are CCSDS draft recommendations (while being developed into Red or Green books).

Only CCSDS participating agencies participate in the consensus process.

CCSDS provides non-binding technical recommendations related to space data handling systems. These recommendations are usually forwarded to ISO for adoption as ISO Standards.

4.4.3 Standards Relevant to ES

CCSDS has four panels for its recommendations in various space data systems in following areas:

Panel 1 - Telemetry, Tracking, and Command;

Panel 2 - Information Interchange Processes;

Panel 3 - Cross Support Operations;

Panel 4 - Radiometric and Orbit Data (currently inactive).

Panel 2 is specifically chartered to investigate Standard Infomation Interchange Processes (SIIP). Table 6 illustrates some standards relevant to ESE.

We should also note that applications such as direct broadcasting and on-board processing may need standards developed by panel 1 and 3. To obtain more information go to the web site: http://www.ccsds.org/

Table 6 List of Relevant ESE Standards

CCSDS	Description	status	relevance
121.0-B-1	Lossless Data Compression. Blue book Issue 1. May 1997	Adopted as ISO 15887:2000	
620.0-B-2.1	Standard Formatted Data Units -- Structure and Construction Rules (with Technical Corrigendum 1). Blue Book. Issue 2. November 1996	Adopted as ISO 12175:1994.	
622.0-B-1	Standard Formatted Data Units - Referencing Environment. Blue Book. Issue 1. May 1997	Adopted as ISO 15888:2000.	
630.0-B-1	Standard Formatted Data Units -- Control Authority Procedures. Blue Book. Issue 1. June 1993	Adopted as ISO 13764:1996	
641.0-B-2	Parameter Value Language Specification (CCSD-0006 and CCSD0008), Issue 2. June 2000.	Adopted as ISO 14961:1997	
643.0-B-1	ASCII Encoded English (CCSD0002).	Adopted as ISO 14962:1997	
644.0-B-2	The Data Description Language EAST Specif-ication (CCSD0010). Blue Book. Issue 2. Novem-ber 2000	adopted as ISO 15889:2000	
647.1-B-1	Data Entity Dictionary Specification Language (DEDSL) - Abstract Syntax (CCSD0011). Blue Book. Issue 1. June 2001	adopted as ISO / DIS	
647.2-B-1	Data Entity Dictionary Specification Language (DEDSL) - PVL Syntax (CCSD0012). Blue Book. Issue 1. June 2001.	adopted as ISO / DIS	
647.3-R-1	Data Entity Dictionary Specification Language	adopted as ISO /	Highly

	(DEDSL) - XML/DTD Syntax (CCSD0013). Red Book. Issue 1. June 2001.	DIS 22643	relevant to ESE
650.0-R-2	Reference Model for an Open Archival Information System (OAIS). Red Book. Issue 2. July 2001	adopted as ISO / DIS 14721.2	
727.0-R-5	CCSDS File Delivery Protocol (CFDP). Red Book. Issue 5. August 2001.	adopted as ISO / DIS 17355	

4.4.4 Current Status

The CCSDS standards are primarily space data systems-oriented, resulting from the international cooperation of space agencies. Therefore these standards are primarily used in space agencies. Some of the standards and concepts could be, and are, used elsewhere.

Table 7 Success and Failure Description

Success Criteria	CCSDS Standard	Comments
Is the development of standards process open and transparent?	Yes	Space agencies are the primary participants and industries doing business in the domain area can also participate by becoming members or affiliate members (there are also non-space organizations participating in Panel 2, information inter-changes. This includes National Archives and National Libraries)
Are the standards developed by this organization freely available - open standards?	Yes	All standards are published and freely available
How much time did it take for the development of the standard- months, years?	Generally it takes 2-4 years for a concept paper to become a recommendation and/or an ISO standard.	For a standards process, this is relatively quick.
Does the standard have software tools support?	Yes	
Vendor support to the standard	It varies from substantial to none, depending on the standard in question.	Many of the CCSDS standards are specific to space data systems and hence do not have broad usage in other industries. . Nevertheless, standards such as CCSDS packets and frames have multiple vendors providing support. Some,like the EAST description language are prim-arily supported by one Vendor and/or a space agency. Others, like a reference model, are not

		implementations and would not have vendor tools.
Relationship with other standards organizations	Good	CCSDS works with many other standards organizations such as ISO and CEOS. ISO has endorsed some CCSDS standards
Has the organization achieved its stated goals?	Yes, in many areas the progress is good. But there is still work in progress in many areas.	
Usage of standards 1. How widely it is used? 2. ESE domain	1. Used in primarily in space agencies worldwide 2. Some	A recent draft standard, the "Reference Model for an Open Archival Information System (OAIS)" is already in wide usage by a vast variety of organizations, including national libraries and traditional archives.
Quality 1. Documentation and description of the standard Multi-language support 2. Quality of the standard test suite	1. Very good 2. Not applicable	
Performance: measured in terms of 1. Usage 2. Adoption 3. Engineering and integration	1. Very good to average 2. Good 3. Very good	This is difficult to generalize as there are many standards and performance data is harder to get
Ease of use	Some of the standards like PVL and SFDU are easy to use and understand.	
Evolution Is there a process for this?	Yes	
Metadata support Interoperability	Very good Excellent	
Are these standards relevant to ESE?	Highly relevant	
NASA participation	Very good	NASA actively participates in CCSDS

4.4.5 NASA Current Involvement

NASA actively participates in all the CCSDS panel activities.

4.5 The U.S. Federal Geographic Data Committee (FGDC)

4.5.1 Description

The Office of Management and Budget (OMB) created the FGDC in 1990, and assigned responsibilities for coordinating geospatial data themes to different federal departments, including the establishment and implementation of data standards for quality, content, and transfer. In 1994, the Clinton administration designated the FGDC as the lead entity to coordinate the National Spatial Data Infrastructure (NSDI), defined as the technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data. In particular, the FGDC was to "develop standards for implementing the NSDI, in consultation and cooperation with state, local, and tribal governments, the private and academic sectors, and … the international community," to promote the use of these standards, and to submit them for consideration as Federal Information Processing Standards. (http://www.fgdc.gov/-standards/refmod97.pdf)

FGDC's work follows two principal overlapping structures: Thematic Subcommittees and Working Groups – as depicted below in Figure 6:

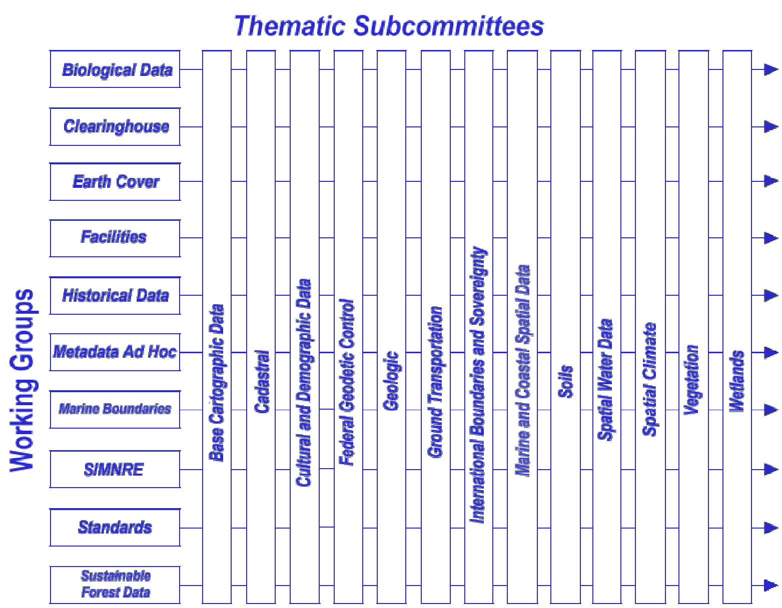

Figure 6 FGDC Thematic Subcommittees & Working Groups

FGDC has adopted four principal means towards building the NSDI:

- *Metadata*: FGDC has developed a content standard for geospatial metadata; and has spurred its widespread adoption through training seminars and seed funding to state and federal data holders. This standard was the basis for ISO-TC211's Metadata standard (19115).
- *Clearinghouse*: FGDC has promoted the notion of a distributed repository ("clearinghouse") of metadata, searchable via a common (Z39.50) network

protocol. Thanks to FGDC's promotional efforts, training support, and seed funding, over 200 clearinghouse "nodes" now provide geospatial metadata in a consistent form, accessible from Z39.50 clients (often Web gateways). This lets users find relevant data listings, and either fetch them directly or contact the appropriate data custodian.

- *Framework*: a collaborative, community-based effort in which public and private organizations within a given geographic area develop, maintain, and integrate a few commonly needed data themes. This lets a variety of users share resources, improve communications, and increase efficiency. -(http://www.fgdc.gov/framework/framework.html)
- *Standards* for implementing the NSDI, in consultation and cooperation with State, local, and tribal governments, the private sector and academic community, and, to the extent feasible, the international community.

4.5.2 Standards Relevant to ES

FGDC's standards list is online at http://www.fgdc.gov/standards/status/text-status.html.

FGDC has several **Endorsed Standards** relevant to Earth Science. The following table highlights a few of these:

FGDC-STD-001-1998	Content Standard for Digital Geospatial Metadata (version 2.0) http://www.fgdc.gov/metadata/contstan.html
FGDC-STD-008-1999	Content Standard for Digital Orthoimagery http://www.fgdc.gov/standards/status/sub3_6.html
FGDC-STD-009-1999	Content Standard for Remote Sensing Swath Data http://www.fgdc.gov/standards/status/sub4_4.html

4.5.3 Standard Work in Progress

FGDC distinguishes draft standards that have completed public review from those still at the draft stage or the proposal stage.

FGDC has several ESE-relevant draft standards that have **completed public review**.

Content Standard for Framework Land Elevation Data	http://www.fgdc.gov/standards/status/sub2_6.html
Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata	http://www.fgdc.gov/standards/status/csdgm_rs_ex.html
U.S. National Grid	http://www.fgdc.gov/standards/status/usng.html

FGDC also has three ES-relevant draft standards in the **draft stage**:

Earth Cover Classification System	http://www.fgdc.gov/standards/status/sub4_6.html
Encoding Standard for Geospatial	http://www.fgdc.gov/standards/status/sub4_5.html

Figure 7 - Information Engineering Four-sided Pyramid

Standards development occurs in 12 steps from initial proposal through adoption. FGDC has adopted these steps from those used in ANSI and ISO processes. These steps are organized into five stages.

Proposal	Steps 1 - 2
Project	Step 3
Draft	Steps 4 - 5
Review	Steps 6 - 11
Final	Step 12

Figure 8 below depicts the process flow through these stages.

Figure 8 FGDC Standards Process Flow Diagram

Criteria for adoption of a standard

As national standards in support of the NSDI, there are several expectations for FGDC Standards (cf. http://www.fgdc.gov/standards/refmod97.pdf):

Within FGDC Scope – governing data or processes to advance data sharing and minimize duplication of effort

Future Focused – should not re-formalize existing solutions, but instead focus on solving future problems, to promote new or enhanced interaction and coordination

Structured – developed and presented according to a structured methodology, so as to be understood and usable

Technology Independent – should not constrain the use of new and emerging technologies

Integrated – with one another and with related standards

Evolving – to accommodate changing technology and institutional mandates, including backward compatibility and known update and maintenance procedures

Supportable – by the geospatial vendor community and by known or emerging technology

Publicly Available: (i) Public notice of availability; (ii) Not developed from copy-righted or proprietary standards; (iii) No copyrights or other limitations on their use or reproduction; (iv) Available electronically whenever possible

Complete and Consistent - complete as to components and methodology; consistent in form and format; responsive to public comments

As for implementation: FGDC Standards apply to and are mandatory for federal agencies. Their use by non-federal and private sector organizations is "encouraged" in order to promote the widest possible use and sharing of data.

FGDC Standards aim to increase interoperability among automated, heterogeneous, autonomously developed geospatial information systems, so as to enable the development of a national digital spatial information resource (the NSDI) with federal, state, local, tribal, and private involvement. This national information resource, linked by criteria and standards, will enable sharing and efficient transfer of spatial data between producers and users. Enhanced coordination will build information partnerships among government institutions and the public and private sectors, avoiding wasteful duplication of effort and ensuring effective and economical management and use of information resources.

4.5.5 NASA Current Involvement

NASA has had the following involvement with FGDC:

- Leads the FGDC Standard Working Group's Imagery subgroup, which is responsible for making imagery related FGDC standards. This subgroup has led two standards developments processes:
    - Digital Geospatial Metadata: Extensions for Remote Sensing Metadata
    - Remote Sensing Swath Data

- Proposed the Global Change Master Directory (GCMD) as a thesaurus for FGDC Clearinghouse metadata collections
- Participates in Clearinghouse, Standards, and Earth Cover working groups
- Coordinated Digital Earth with NSDI
- Leads new Geospatial Applications & Interoperability (GAI) working group

4.6 The Internet Engineering Task Force (IETF)

4.6.1 Description

The Internet Engineering Task Force (IETF) (cf. http://www.ietf.org/rfc/rfc2026.-txt) is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of Internet architecture and the smooth operation of the Internet. It is open to any interested individual. IETF working groups are grouped into Areas, and managed by Area Directors, or ADs. The ADs are members of the Internet Engineering Steering Group (IESG). Providing architectural oversight is the Internet Architecture Board (IAB). The IAB also adjudicates appeals when someone complains that the IESG has failed. The Internet Society charters the IAB and IESG for these purposes. The General Area Director also serves as the chair of the IESG and of the IETF, and is an *ex-officio* member of the IAB.

4.6.2 Standards Relevant to ES

IETF RFCs define such ubiquitous infrastructure as TCP/IP, POP and SMTP email transfer, FTP file transfer, etc.

IETF RFCs number in the thousands, with widely varying scope. Search terms such as “geographic,” “Earth,” and “multimedia,” will retrieve relevant RFCs; as will terms such as “XML,” “Web,” and “IPv6”.

4.6.3 Standard Work in Progress

A few draft IETF standards touch on geospatial / Earth science topics – for instance:

- Geographic registration of HTML documents
- Geographic extensions for HTTP transactions
- Enhancing TCP Over Satellite Channels using Standard Mechanisms
- Definition of the DNS Geographic Location (GL) Resource Record

4.6.4 Standard Processes

(http://www.ietf.org/rfc/rfc2026.txt)

The Internet Standards Process is concerned with all protocols, procedures, and conventions that are used in or by the Internet, whether or not they are part of the

TCP/IP protocol suite. In the case of protocols developed and/or standardized by non-Internet organizations, however, the Internet Standards Process normally applies to the application of the protocol or procedure in the Internet context, not to the specification of the protocol itself.

In principle, the process of creating an Internet Standard is straightforward: a specification undergoes a period of development and several iterations of review by the Internet community and revision based upon experience, is adopted as a Standard by the appropriate body (see below), and is published. In practice, the process is more complicated, due to (1) the difficulty of creating specifications of high technical quality; (2) the need to consider the interests of all of the affected parties; (3) the importance of establishing widespread community consensus; and (4) the difficulty of evaluating the utility of a particular specification for the Internet community.

Requests for Comments (RFCs)

Each distinct version of an Internet standards-related specification is published as part of the "Request for Comments" (RFC) document series. This archival series is the official publication channel for Internet standards documents and other publications of the IESG, IAB, and Internet community. RFCs can be obtained from a number of Internet hosts using anonymous FTP, gopher, World Wide Web, and other Internet document-retrieval systems.

Some RFCs document Internet Standards: these form the “STD” subseries of the RFC series. Others describe “Best Current Practices,” and form the “BCP” subseries. (RFCs not intended as standards are labeled "Experimental" or "Informational.") Specifications subject to the Internet Standards Process fall into one of two categories: a Technical Specification (TS) describes a protocol, service, procedure, convention, or format, whereas an Applicability Statement (AS) specifies how one or more TSs must / should / may be applied to support a particular Internet capability.

During the development of a specification, draft versions of the document are made available for review in the IETF's "Internet-Drafts" directory, which is replicated on a number of Internet hosts. Internet-Drafts have no formal status, and are subject to change or removal at any time An Internet-Draft that is published as an RFC, or that has remained unchanged in the Internet-Drafts directory for more than six months without being recommended for publication as an RFC, is removed from the Internet-Drafts directory. At any time, an Internet-Draft may be replaced by a more recent version of the same specification, restarting the six-month timeout period.

Specifications that are intended to become Internet Standards evolve through a set of maturity levels -- "Proposed Standard," "Draft Standard," and "Standard" – collectively known as the "standards track."

- A *Proposed Standard* specification has undergone enough verification to be con-sidered stable, and has enough community interest to be considered valuable. Implementation or operational experience is desirable, but not required.

- A *Draft Standard* requires at least two independent and interoperable implementations from different code bases, and full-scale operational experience. A Draft Standard is normally considered to be a final specification.
- An *Internet Standard* is one that has seen significant implementation and successful operational experience; it has a high degree of technical maturity and is believed to provide significant benefit to the Internet community.

Criteria for adoption of a standard

In general, an Internet Standard is a specification that is stable and well-understood, is technically competent, has multiple, independent, and interoperable implementations with substantial operational experience, enjoys significant public support, and is recognizably useful in some or all parts of the Internet.

The goals of the Internet Standards Process are:

- Technical excellence;
- Prior implementation and testing;
- Clear, concise, and easily understood documentation;
- Openness and fairness; and
- Timeliness.

The goal of technical competence, the requirement for prior implementation and testing, and the need to allow all interested parties to comment all require significant time and effort. On the other hand, today's rapid development of networking technology demands timely development of standards. The Internet Standards Process is intended to balance these conflicting goals. The process is believed to be as short and simple as possible without sacrificing technical excellence, thorough testing before adoption of a standard, or openness and fairness.

From its inception, the Internet has been, and is expected to remain, an evolving system whose participants regularly factor new requirements and technology into its design and implementation. Users of the Internet and providers of the equipment, software, and services that support it should anticipate and embrace this evolution as a major tenet of Internet philosophy.

A. APPENDIX A: LIST OF ACRONYMS

Acronym	Description
AD	Area Directors
AS	Applicability Statement *or* OGC Abstract Specification
BCP	Best Current Practices
CAN	
CCSDS	Consultative Committee for Space Data Systems
DAAC	Distributed Active Archive Center
DAN	
DDN	
DIF	Directory Interchange Format
DP	Discussion Paper
ECS	EOSDIS Core System
EO	Earth Observation
EOSDIS	NASA Earth Observing System Data and Information System
ES	Earth Science
ESE	Earth Science Enterprise
ESIP	NASA Earth Science Information Partners
ESTO	NASA Earth Science Technology Office
FDIS	Final Draft International Standard
FGDC	Federal Geographic Data Committee
GAD	General Area Director
GCMD	Global Change Master Directory
GDAAC	Goddard Distributed Active Archive Center
GML	Geography Markup Language
HDF	Hierarchical Data Format
IAB	Internet Architecture Board
IEC	International Electrotechnical Commission (often seen as ISO/IEC)
IESG	Internet Engineering Steering Group
IP	Internet Protocol *or* OGC Interoperability Program
IPv6	Internet Protocol version 6
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
IETF	Internet Engineering Task Force
IMS	Information Management System
ISO	International Organization for Standardization
ISO TC	International Organization for Standardization Technical Committee
JESST	Java Earth Science Search Tool
MathML	Mathematical Markup Language
Mercury	
NCSA	National Center for Supercomputing Applications
NSDI	National Spatial Data Infrastructure
ODL	Object Description Language
OGC	Open GIS Consortium
OMB	Office of Management and Budget
OMG	Object Management Group
PC	OGC Planning Committee
POP	Post Office Protocol *or* Point Of Presence
RFC	Request For Comment
RFP	Request For Proposal
RP	Recommendation Paper
SEEDS	Strategic Evolution of ESE Data Systems

SIG	Special Interest Group
SMTP	Simple Mail Transfer Protocol
SP	OGC Specification Program
SVG	Scalable Vector Graphics
TAG	Technology Architecture Group
TC	Technical Committee
TC 211	ISO Technical Committee #211, Geographic Information / Geomatics
TCP/IP	Transmission Control Protocol/Internet Protocol
TS	Technical Specification
TSDIS	TRMM Science Data and Information System
XML	Extensible Markup Language
WCS	OpenGIS Web Coverage Service
WFS	OpenGIS Web Feature Service
WMS	OpenGIS Web Map Service
WG	Working Group
W3C	World Wide Web Consortium
Z39.50	Also known as ISO 23950: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification."

B. APPENDIX B: LIST OF URLS

URL	Description
http://www.ccsds.org/	CCSDS web site
http://esipfed.net/committees/interop/swil.html	Federation swil tools and documentation
http://www.fgdc.gov/standards/refmod97.pdf	FGDC Standards Reference Model
http://www.fgdc.gov/framework/framework.html	FGDC framework
http://www.fgdc.gov/standards/status/textstatus.html	FGDS's standard list
http://gcmd.gsfc.nasa.gov	GCMD web site
http://www.ietf.org/rfc/rfc2026.txt	IETF standard process
http://www.ietf.org/glossary.html#IETF	IETF glossary
http://www.fgdc.gov/nsdi/nsdi.html	NSDI
http://www.opengis.org	OGC web site
http://www.fgdc.gov/metadata/contstan.html	Content standard for Digital Metadata
http://mcmcweb.er.usgs.gov/sdts/	Spatial Data transfer Protocol
http://www-v0ims.gsfc.nasa.gov/v0ims/DOCUMENTATION/documentation.html	V0 tools and documentation
http://www.w3.org/	W3C web site